

The Delphion Integrated View

Get Now: ☒ PDF | More choices...

Tools: Add to Work File:

View: INPADOC | Jump to: Go to: Derwent

Title: **JP11232383A2: METHOD AND DEVICE FOR TABLE PROCESSING**

Derwent Title: Table processing method for character recognition of front area in image - involves choosing rectangle frame of position corresponding to specific frame distribution in image extracted frame positional information as specific rectangle frame [Derwent Record]

Country: JP Japan
Kind: A

Inventor: OOKUMA YOSHINORI;
ITO KOJI;

Assignee: OKI ELECTRIC IND CO LTD
News, Profiles, Stocks and More about this company

Published / Filed: 1999-08-27 / 1998-02-16

Application Number: JP1998000033260

IPC Code: G06K 9/20; G06F 19/00;

Priority Number: 1998-02-16 JP1998199833260


Abstract: PROBLEM TO BE SOLVED: To attain reduction of time needed for table processing.

SOLUTION: A table processor 10 is equipped with a table format/frame arrangement storage part 12 that associates a title with a table format and stores them as table format information, associates an item with a data frame every table format and stores them as frame arrangement information, a table area/frame position extraction part 16 for extracting positions of a table area and each rectangular frame from an image picture, a title character string area extraction part 18 for extracting a title character string area from a non- table area, a table format/frame arrangement selection part 20 that selects the table format corresponding to an extracted title out of table format information and selects the arrangement of a data frame corresponding to the specified item as a specified frame arrangement out of the frame arrangement information, and a table processing part 22 that selects a rectangular frame at a position corresponding to the specified frame arrangement as a specified rectangular frame and extracts a specified data character string area from within the specified rectangular frame.

COPYRIGHT: (C)1999,JPO

Family: None

Forward References: **Go to Result Set:** Forward references (1)

PDF	Patent	Pub.Date	Inventor	Assignee	Title
	US6709560	2004-03-23	Andelman; Marc D.	Biosource, Inc.	Charge barrier flow-thr

<https://www.delphion.com/details?pn=JP11232383A2>

6/7/2004

BEST AVAILABLE COPY

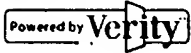
capacitor

Other Abstract
Info:

DERABS G1999-533889 DERABS G1999-533889



Nominate this for the Gall



© 1997-2004 Thomson

[Research Subscriptions](#) | [Privacy Policy](#) | [Terms & Conditions](#) | [Site Map](#) | [Contact Us](#) | [Feedback](#)

BEST AVAILABLE COPY



(19)

(11) Publication number: **11232383 A**

Generated Document.

PATENT ABSTRACTS OF JAPAN

(21) Application number: 10033260

(51) Intl. Cl.: G06K 9/20 G06F 19/00

(22) Application date: 16.02.98

(30) Priority:

(43) Date of application publication: 27.08.99

(84) Designated contracting states:

(71) Applicant: OKI ELECTRIC IND CO LTD

(72) Inventor: **OOKUMA YOSHINORI**
ITO KOJI

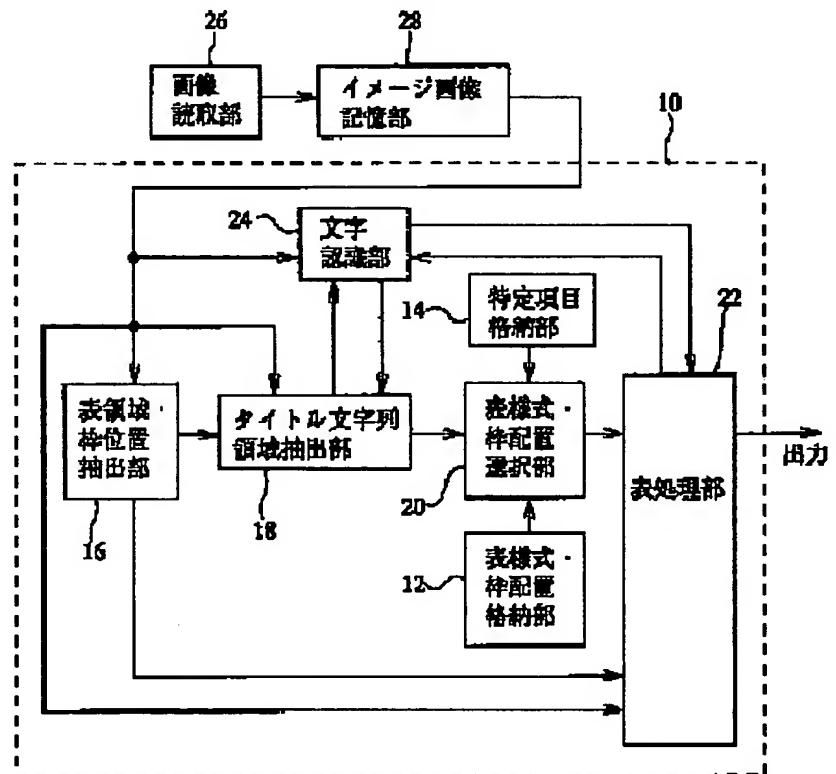
(74) Representative:

(54) METHOD AND DEVICE FOR TABLE PROCESSING

(57) Abstract:

PROBLEM TO BE SOLVED: To attain reduction of time needed for table processing.

SOLUTION: A table processor 10 is equipped with a table format/frame arrangement storage part 12 that associates a title with a table format and stores them as table format information, associates an item with a data frame every table format and stores them as frame arrangement information, a table area/frame position extraction part 16 for extracting positions of a table area and each rectangular frame from an image picture, a title character string area extraction part 18 for extracting a title character string area from a non- table area, a table format/frame arrangement selection part 20 that selects the table format corresponding to an extracted title out of table format information and selects the arrangement of a data frame corresponding to the specified item as a specified frame arrangement out of the frame arrangement information, and a table processing part 22 that selects a rectangular frame at a position corresponding to the specified frame arrangement as a specified rectangular frame and extracts a specified data character string area from within the specified



rectangular frame.

COPYRIGHT: (C)1999,JPO

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-232383

(43) 公開日 平成11年(1999) 8月27日

(51) IntCl.⁸

G 0 6 K 9/20

G 0 6 F 19/00

識別記号

3 4 0

F I

G 0 6 K 9/20

G 0 6 F 15/22

3 4 0 L

Z

審査請求 未請求 請求項の数15 O L (全 26 頁)

(21) 出願番号

特願平10-33260

(22) 出願日

平成10年(1998) 2月16日

(71) 出願人 000000295

沖電気工業株式会社

東京都港区虎ノ門1丁目7番12号

(72) 発明者 大熊 好憲

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

(72) 発明者 伊東 晃治

東京都港区虎ノ門1丁目7番12号 沖電気
工業株式会社内

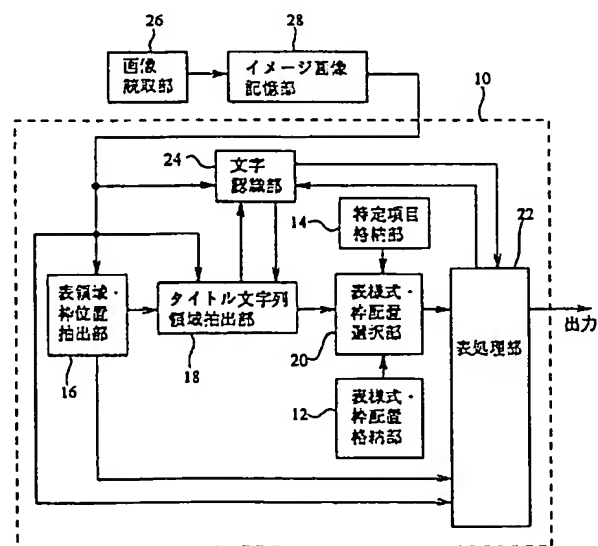
(74) 代理人 弁理士 大垣 孝

(54) 【発明の名称】 表処理方法および表処理装置

(57) 【要約】

【課題】 表処理に要する時間の短縮を図る。

【解決手段】 タイトルと表様式とを対応づけて表様式情報として格納し、表様式毎に項目とデータ枠とを対応づけて枠配置情報として格納しておく表様式・枠配置格納部12と、イメージ画像から表領域および各矩形枠の位置を抽出する表領域・枠位置抽出部16と、非表領域からタイトル文字列領域を抽出するタイトル文字列領域抽出部18と、表様式情報の中から抽出タイトルに対応する表様式を選択し、特定項目に対応するデータ枠の配置を枠配置情報の中から特定枠配置として選択する表様式・枠配置選択部20と、特定枠配置に該当する位置の矩形枠を特定矩形枠として選択し、当該特定矩形枠内から特定データ文字列領域を抽出する表処理部22とを具える。



10: 表処理装置

【特許請求の範囲】

【請求項1】 表の特定項目に対応するデータ（以下、「特定データ」とも称する。）を当該表のイメージ画像から抽出するにあたり、

表様式・枠配置格納部に、表のタイトルと表様式とを対応づけて表様式情報として予め格納しておくと共に、前記表様式毎に、当該表様式に属する項目と当該項目に対応するデータが記載されたデータ枠の表中での配置とを対応づけて枠配置情報として予め格納しておき、

（a）前記イメージ画像から、表領域を抽出すると共に、該表領域を構成する各矩形枠の位置を枠位置情報として抽出する処理と、

（b）前記イメージ画像のうちの非表領域から、前記表のタイトルを抽出タイトルとして抽出する処理と、

（c）前記表様式・枠配置格納部から前記表様式情報を読み出して、当該表様式情報の中から前記抽出タイトルに対応する表様式を選択表様式として選択する処理と、

（d）前記選択表様式に対応する前記枠配置情報を前記表様式・枠配置格納部から読み出して、当該枠配置情報の中から前記特定項目に対応するデータ枠の表中での配置を特定枠配置として選択する処理と、

（e）前記（a）の処理で抽出された前記枠位置情報の中から前記イメージ画像において前記特定枠配置に該当する位置の矩形枠を特定矩形枠として選択する処理と、

（f）前記特定矩形枠内から前記特定データを抽出する処理とを含むことを特徴とする表処理方法。

【請求項2】 請求項1に記載の表処理方法において、前記（a）の処理は、

前記イメージ画像を複数の方向にそれぞれ走査して、前記方向毎に、走査線毎の累積黒画素数を検出して周辺分布を作成する処理と、

前記周辺分布に基づいて、前記イメージ画像から縦罫線および横罫線をそれぞれ抽出する処理と、

抽出された前記縦罫線および前記横罫線に囲まれた前記矩形枠をそれぞれ抽出し、抽出された前記矩形枠の位置を前記枠位置情報とする処理と、

各前記矩形枠でもって構成される前記表領域を抽出する処理とを含むことを特徴とする表処理方法。

【請求項3】 請求項1に記載の表処理方法において、前記（b）の処理は、

前記非表領域からタイトル文字列領域を抽出する処理と、

前記イメージ画像のうちの前記タイトル文字列領域について文字認識処理を行って、前記抽出タイトルを抽出する処理とを含むことを特徴とする表処理方法。

【請求項4】 請求項1に記載の表処理方法において、前記（f）の処理は、

前記特定矩形枠内から特定データ文字列領域を抽出する処理と、

前記特定データ文字列領域について文字認識処理を行っ

て、前記特定データを抽出する処理とを含むことを特徴とする表処理方法。

【請求項5】 表の特定項目に対応するデータ（以下、「特定データ」とも称する。）を当該表のイメージ画像から抽出する表処理装置において、

表様式・枠配置格納部と、特定項目格納部と、表領域・枠位置抽出部と、タイトル文字列領域抽出部と、表様式・枠配置選択部と、表処理部と、文字認識部とを具え、前記表様式・枠配置格納部は、表のタイトルと表様式とを対応づけて表様式情報として格納しておくと共に、前記表様式毎に、当該表様式に属する項目と当該項目に対応するデータが記載されたデータ枠の表中での配置とを対応づけて枠配置情報として格納しておくための構成を有し、

前記特定項目格納部は、前記特定項目を格納しておくための構成を有し、

前記表領域・枠位置抽出部は、前記イメージ画像から、表領域を抽出すると共に、該表領域を構成する各矩形枠の位置を枠位置情報として抽出する構成を有し、

前記タイトル文字列領域抽出部は、前記イメージ画像のうちの非表領域からタイトル文字列領域を抽出し、当該タイトル文字列領域を前記文字認識部に転送し、前記文字認識部から抽出タイトルを取得する構成を有し、

前記表様式・枠配置選択部は、前記タイトル文字列領域抽出部から前記抽出タイトルを取得すると共に、前記表様式・枠配置格納部から前記表様式情報を読み出して、当該表様式情報の中から当該抽出タイトルに対応する表様式を選択表様式として選択し、かつ、当該選択表様式に対応する前記枠配置情報を前記表様式・枠配置格納部から読み出すと共に、前記特定項目格納部から前記特定項目を読み出して、当該特定項目に対応するデータ枠の表中での配置を当該枠配置情報の中から特定枠配置として選択する構成を有し、

前記表処理部は、前記表様式・特定枠配置選択部から前記特定枠配置を取得すると共に、前記表領域・枠位置抽出部から前記枠位置情報を取得して、当該特定枠配置に該当する位置の矩形枠を当該枠位置情報の中から特定矩形枠として選択し、かつ、当該特定矩形枠内から特定データ文字列領域を抽出し、当該特定データ文字列領域を前記文字認識部に転送し、前記文字認識部から前記特定データを取得する構成を有し、

前記文字認識部は、前記タイトル文字列領域抽出部から前記タイトル文字列領域を取得し、前記イメージ画像のうちの当該タイトル文字列領域について文字認識を行って、前記抽出タイトルを抽出し、かつ、前記表処理部から前記特定データ文字列領域を取得して、前記イメージ画像のうちの前記特定データ文字列領域について文字認識を行って、前記特定データを抽出する構成を有することを特徴とする表処理装置。

【請求項6】 請求項5に記載の表処理装置において、

前記表領域・枠位置抽出部は、

前記イメージ画像を複数の方向にそれぞれ走査して、前記方向毎に、走査線毎の累積黒画素数を検出して周辺分布を作成する周辺分布作成部と、

前記周辺分布に基づいて、前記イメージ画像から縦罫線および横罫線をそれぞれ抽出する罫線抽出部と、抽出された前記縦罫線および前記横罫線に囲まれた前記矩形枠をそれぞれ抽出し、抽出された前記矩形枠の位置を前記枠位置情報とする矩形枠抽出部と、各前記矩形枠をもって構成される前記表領域を表領域抽出部とを含むことを特徴とする表処理装置。

【請求項7】 表の特定項目に対応するデータ（以下、「特定データ」とも称する。）を当該表のイメージ画像から抽出するにあたり、

表様式格納部に、表のタイトルと表様式とを対応づけて表様式情報として予め格納しておくと共に、前記表様式毎に、当該表様式に属する項目の記載された項目枠と当該項目に対応するデータが記載されたデータ枠との配置関係を枠配置関係情報として予め格納しておき、

（a）前記イメージ画像から、表領域を抽出すると共に、該表領域を構成する各矩形枠の位置を枠位置情報として抽出する処理と、

（b）前記イメージ画像のうちの非表領域から前記表のタイトルを抽出タイトルとして抽出する処理と、

（c）前記表様式格納部から前記表様式情報を読み出して、当該表様式情報の中から前記抽出タイトルに対応する表様式を選択表様式として選択し、かつ、前記表様式格納部から当該選択表様式に対応する枠配置関係情報を選択枠配置関係情報として読み出す処理と、

（d）前記（a）の処理で抽出された前記枠位置情報の示す前記矩形枠の中から項目枠を選択する処理と、

（e）前記項目枠の各々に記載された項目をそれぞれ抽出項目として抽出する処理と、

（f）前記抽出項目として前記特定項目が記載された項目枠を前記項目枠の中から特定項目枠として選択する処理と、

（g）前記特定項目枠の位置に対して前記選択枠配置関係情報の示す位置にある矩形枠を前記イメージ画像から特定矩形枠として抽出する処理と、

（h）前記特定矩形枠内から前記特定データを抽出する処理とを含むことを特徴とする表処理方法。

【請求項8】 請求項7に記載の表処理方法において、前記（a）の処理は、

前記イメージ画像を複数の方向にそれぞれ走査して、前記方向毎に、走査線毎の累積黒画素数を検出して周辺分布を作成する処理と、

前記周辺分布に基づいて、前記イメージ画像から縦罫線および横罫線をそれぞれ抽出する処理と、

抽出された前記縦罫線および前記横罫線に囲まれた前記矩形枠をそれぞれ抽出し、抽出された前記矩形枠の位置

を前記枠位置情報とする処理と、

各前記矩形枠をもって構成される前記表領域を抽出する処理とを含むことを特徴とする表処理方法。

【請求項9】 請求項7に記載の表処理方法において、前記（b）の処理は、

前記非表領域からタイトル文字列領域を抽出する処理と、

前記タイトル文字列領域について文字認識処理を行って、前記抽出タイトルを抽出する処理とを含むことを特徴とする表処理方法。

【請求項10】 請求項7に記載の表処理方法において、前記（d）の処理は、

前記矩形枠の面積を求める処理と、

前記矩形枠のうち、一定面積以下の面積を有する矩形枠を前記項目枠として選択する処理とを含むことを特徴とする表処理方法。

【請求項11】 請求項7に記載の表処理方法において、前記（e）の処理は、

前記項目枠から項目文字列領域を抽出する処理と、

前記項目文字列領域について文字認識を行って、前記抽出項目を抽出する処理とを含むことを特徴とする表処理方法。

【請求項12】 請求項7に記載の表処理方法において、前記（h）の処理は、

前記特定矩形枠内から特定データ文字列領域を抽出する処理と、

前記特定データ文字列領域について文字認識処理を行って、前記特定データを抽出する処理とを含むことを特徴とする表処理方法。

【請求項13】 表の特定項目に対応するデータ（以下、「特定データ」とも称する。）を当該表のイメージ画像から抽出する表処理装置において、

表様式格納部と、特定項目格納部と、表領域・枠位置抽出部と、タイトル文字列領域抽出部と、表様式選択部と、表処理部と、文字認識部とを具え、

前記表様式格納部は、表のタイトルと表様式とを対応づけて表様式情報として格納しておくと共に、前記表様式毎に、当該表様式に属する項目の記載された項目枠と当該項目に対応するデータが記載されたデータ枠との配置関係を枠配置関係情報として格納しておくための構成を有し、

前記特定項目格納部は、前記特定項目を格納しておくための構成を有し、

前記表領域・枠位置抽出部は、前記イメージ画像から、表領域を抽出すると共に、該表領域を構成する各矩形枠の位置を枠位置情報として抽出する構成を有し、

前記タイトル文字列領域抽出部は、前記イメージ画像のうちの非表領域からタイトル文字列領域を抽出し、当該タイトル文字列領域を前記文字認識部に転送し、前記文字認識部から抽出タイトルを取得する構成を有し、

前記表様式選択部は、前記タイトル文字列領域抽出部から前記抽出タイトルを取得すると共に、前記表様式格納部から前記表様式情報を読み出して、当該表様式情報の中から当該抽出タイトルに対応する表様式を選択表様式として選択し、かつ、前記表様式格納部から該選択表様式に対応する枠配置関係情報を選択枠配置関係情報として読み出す構成を有し、

前記表処理部は、前記表領域・枠位置抽出部から前記枠位置情報を取得して、当該枠位置情報の示す前記矩形枠の中から項目枠を選択し、前記項目枠の各々から項目文字列領域を抽出し、当該項目文字列領域を前記文字認識部に転送し、前記文字認識部から抽出項目を取得し、前記特定項目格納部から前記特定項目を読み出して、前記項目枠の中から当該特定項目が前記抽出項目として記載された項目枠を特定項目枠として選択し、前記表様式選択部から前記選択枠配置関係情報を取得して、前記特定項目枠の位置に対して当該選択枠配置関係情報の示す位置にある矩形枠を前記イメージ画像から特定矩形枠として抽出し、かつ、前記特定矩形枠内から特定データ文字列領域を抽出し、当該特定データ文字列領域を前記文字認識部に転送し、前記文字認識部から前記特定データを取得する構成を有し、

前記文字認識部は、前記イメージ画像のうちの前記タイトル文字列領域について文字認識を行って前記抽出タイトルを抽出し、前記イメージ画像のうちの前記項目文字列領域について文字認識を行って前記項目を抽出し、かつ、前記イメージ画像のうちの前記特定データ文字列領域について文字認識を行って前記特定データを抽出する構成を有してなることを特徴とする表処理装置。

【請求項14】 請求項13に記載の表処理装置において、表領域・枠位置抽出部は、前記イメージ画像を複数の方向にそれぞれ走査して、前記方向毎に、走査線毎の累積黒画素数を検出して周辺分布を作成する周辺分布作成部と、前記周辺分布に基づいて、前記イメージ画像から縦罫線および横罫線をそれぞれ抽出する罫線抽出部と、抽出された前記縦罫線および前記横罫線に囲まれた前記矩形枠をそれぞれ抽出し、抽出された前記矩形枠の位置を前記枠位置情報とする矩形枠抽出部と、各前記矩形枠でもって構成される前記表領域を抽出する表領域抽出部とを具えてなることを特徴とする表処理装置。

【請求項15】 請求項13に記載の表処理装置において、前記表処理部は、前記表領域から前記項目枠を選択するにあたり、前記矩形枠の面積を求め、前記矩形枠のうち、一定面積以下の面積を有する矩形枠を前記項目枠として選択する構成を有することを特徴とする表処理装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、画像処理により、イメージ画像中の表領域の文字認識を行う表処理装置および方法に関する。

【0002】

【従来の技術】従来の表処理装置の一例が、文献：「特開平3-74728号公報」に開示されている。この文献に開示の技術によれば、先ず、イメージ画像の表領域の内外にある全ての文字ボタンについて文字認識を行う。次に、表領域の外にあるタイトル文字を上位キーワードとし、かつ、表領域中にある文字を下位キーワードとする。そして、上位および下位キーワードに基づいて、表中の必要とするデータを階層的に抽出する。

【0003】

【発明が解決しようとする課題】しかしながら、上記の従来例においては、表領域中の特定の項目のデータのみを抽出する表処理を行う場合も、表領域中の全ての文字ボタンを切り出して、全ての文字ボタンについて文字認識を行っていた。すなわち、従来は、表領域中の必要としない情報についても文字認識を行っていた。その結果、従来例の方法では、表処理にあたり、必要とするデータの多少に関わらず、表領域中の全ての文字ボタンの文字認識を行う場合と同じ時間がかかるという問題点があった。

【0004】このため、表処理に要する時間の短縮が図れる表処理方法および装置の出現が望まれていた。

【0005】

【課題を解決するための手段】この出願にかかる発明者は、種々の検討および研究を重ねた結果、表様式によって、特定の項目のデータ（特定データ）が記載された欄（矩形枠）が、表領域中の特定の位置に配置されている傾向があることに着目した。そして、この発明者は、この表様式を利用すれば、表領域の全てについて文字認識を行わなくとも、表領域の一部分について選択的に文字認識を行えば、表領域中の必要とするデータを従来よりも短い時間で抽出できると考えた。

【0006】さらに、この発明者は、表様式を判定するために、表のタイトルを利用することに想到した。そして、このタイトルと表様式とを対応づけて登録しておけば、タイトルのみを文字認識することにより、表様式を判定できることに想到した。

【0007】（第1の表処理方法）そこで、この発明の第1の表処理方法によれば、表の特定項目に対応するデータ（以下、「特定データ」とも称する。）を当該表のイメージ画像から抽出するにあたり、先ず、表様式・枠配置格納部に、表のタイトルと表様式とを対応づけて表様式情報として予め格納しておく。さらに、この表様式・枠配置格納部に、表様式毎に、当該表様式に属する項目と当該項目に対応するデータが記載されたデータ枠の表領域中での配置とを対応づけて枠配置情報として予め

格納しておく。

【0008】そして、第1の表処理方法では、(a)イメージ画像から、表領域を抽出すると共に、この表領域を構成する各矩形枠の位置を枠位置情報として抽出する。次に、(b)イメージ画像のうちの非表領域から、表のタイトルを抽出タイトルとして抽出する。さらに、(c)表様式・枠配置格納部から表様式情報を読み出して、当該表様式情報の中から抽出タイトルに対応する表様式を選択表様式として選択する。続いて、(d)選択表様式に対応する枠配置情報を表様式・枠配置格納部から読み出して、当該枠配置情報の中から特定項目に対応するデータ枠の表中での配置を特定枠配置として選択する。そして、(e)(a)の処理で抽出された枠位置情報の中からイメージ画像において特定枠配置に該当する位置の矩形枠を特定矩形枠として選択する。最後に、(f)特定矩形枠内から特定データを抽出する。

【0009】そして、第1の表処理方法によれば、表領域の全てについて文字認識を行わなくとも、特定データを選択的に抽出できる。従って、表領域の全てについて文字認識を行う場合に要する処理時間に比べて、表処理に要する処理時間の短縮を図ることができる。次の段落において、特定データを選択的に抽出できる理由について説明する。

【0010】第1の表処理方法においては、タイトルと表様式とを対応づけて予め登録しておく。このように登録しておけば、イメージ画像から抽出したタイトル(抽出タイトル)を用いて、表様式を選択することができる。さらに、この処理方法では、各表様式について、当該表様式に属する項目毎のデータ枠の配置を予め登録しておく。このように登録しておけば、特定項目を指定することにより、選択された表様式におけるその特定項目に対応するデータ枠の配置(特定枠配置)を決定することができる。その結果、イメージ画像においてこの特定枠配置に該当する位置にある矩形枠(特定矩形枠)の領域を決定することができる。すなわち、表領域について文字認識を行う前に、イメージ画像中の特定データが記載された特定矩形枠の領域を決定することができる。従って、その特定矩形枠の領域について選択的に文字認識をすることができる。その結果、第1の処理方法によれば、表領域から特定データを選択的に抽出することができる。すなわち、表領域の全てについて文字認識を行わなくとも、特定データを選択的に抽出することができる。このため、表領域の全てについて文字認識を行ってから特定データを抽出する場合に要する処理時間に比べて、この処理方法では、特定データの抽出に要する処理時間の短縮を図ることができる。すなわち、表処理に要する時間の短縮を図ることができる。

【0011】次に、第1の表処理方法についてより詳細に説明する。

【0012】第1の表処理方法では、表処理の実行に先

立ち、前述のように、タイトルと表様式とを対応づけて表様式情報として登録しておく。この表様式情報は、記憶部である表様式・枠配置格納部に読み出し自在に格納しておく。また、各表様式毎に、各項目のデータ枠の表における配置を枠配置情報として登録しておく。そして、この枠配置情報も、表様式・枠配置格納部に読み出し自在に格納しておく。

【0013】そして、この処理方法では、(a)まず、イメージ画像から表領域を抽出する。この際、表を構成する矩形枠の位置も枠位置情報として抽出する。

【0014】次に、(b)そのイメージ画像のうちの非表領域、すなわち、表領域の外側の領域から表のタイトルを抽出する。

【0015】次に、この処理方法では、(c)抽出されたタイトル(抽出タイトル)を用いて、表様式を決定する。表様式の決定にあたっては、まず、表様式・枠配置格納部から表様式情報を読み出す。次に、抽出タイトルと、表様式情報に含まれるタイトルとを順次に照合する。照合の結果、抽出タイトルと一致したタイトルを選択する。そして、表様式情報から、この選択されたタイトルに対応づけられて登録されている表様式を、選択表様式として選択する。

【0016】続いて、この処理方法では、(d)選択表様式における特定枠配置を決定する。特定枠配置の決定にあたっては、まず、特定項目と、選択表様式に含まれる項目とを照合する。特定項目は、例えば特定項目格納部に格納しておくが良い。そして、照合の結果、特定項目と一致する項目に対応する特定矩形枠の表領域中での配置(特定枠配置)を枠配置情報の中から選択する。

【0017】ところで、上記の(d)の処理で求めた特定枠配置は、特定のデータ枠の表様式における配置を示しているにすぎない。一方、イメージ画像における各矩形枠の位置は、上記の(a)の処理において、枠位置情報として求めている。そこで、(e)枠位置情報の中から、イメージ画像において特定枠配置に該当する位置の矩形枠を特定矩形枠として選択する。

【0018】最後に、この処理方法では、(f)特定矩形枠内から特定データを抽出する。

【0019】また、第1の表処理方法において、好ましくは、枠配置情報は、各データ枠の配置を、表における行番号および列番号でそれぞれ表すのが良い。行番号および列番号を用いれば、表における枠の配置を容易かつ正確に表すことができる。

【0020】また、第1の表処理方法の実施にあたり、(a)の処理は、イメージ画像を複数の方向にそれぞれ走査して、方向毎に、走査線毎の累積黒画素数を検出して周辺分布を作成する処理と、周辺分布に基づいて、イメージ画像から縦罫線および横罫線をそれぞれ抽出する処理と、抽出された縦罫線および横罫線に囲まれた矩形枠をそれぞれ抽出し、抽出された矩形枠の位置を枠位置

情報とする処理と、各矩形枠をもって構成される表領域を抽出する処理とを含むのが好適である。

【0021】さらに、この表処理方法の実施にあたり、好ましくは、枠位置情報は、矩形枠の位置と、表領域における当該矩形枠の行番号および列番号とを対応づけてあると良い。

【0022】また、第1の表処理方法の実施にあたり、(b)の処理は、非表領域からタイトル文字列領域を抽出する処理と、イメージ画像のうちのタイトル文字列領域について文字認識処理を行って、抽出タイトルを抽出する処理とを含むことが望ましい。

【0023】また、第1の表処理方法の実施にあたり、(f)の処理は、特定矩形枠内から文字列領域を抽出する処理と、イメージ画像のうちの文字列領域について文字認識処理を行って、特定データを抽出する処理とを含むことが望ましい。

【0024】(第1の表処理装置)この発明の第1の表処理装置によれば、表の特定項目に対応するデータ(以下、「特定データ」とも称する。)を当該表のイメージ画像から抽出する表処理装置において、表様式・枠配置格納部と、特定項目格納部と、表領域・枠位置抽出部と、タイトル文字列領域抽出部と、表様式・枠配置選択部と、表処理部と、文字認識部とを具えている。

【0025】この表様式・枠配置格納部は、表のタイトルと表様式とを対応づけて表様式情報として格納しておくと共に、表様式毎に、当該表様式に属する項目と当該項目に対応するデータが記載されたデータ枠の表中での配置とを対応づけて枠配置情報として格納しておくための構成を有する。

【0026】また、特定項目格納部は、特定項目を格納しておくための構成を有する。

【0027】また、表領域・枠位置抽出部は、イメージ画像から、表領域を抽出すると共に、該表領域を構成する各矩形枠の位置を枠位置情報として抽出する構成を有する。

【0028】また、タイトル文字列領域抽出部は、イメージ画像のうちの非表領域からタイトル文字列領域を抽出し、当該タイトル文字列領域を文字認識部に転送し、文字認識部から抽出タイトルを取得する構成を有する。

【0029】また、表様式・枠配置選択部は、先ず、タイトル文字列領域抽出部から抽出タイトルを取得する。さらに、表様式・枠配置選択部は、表様式・枠配置格納部から表様式情報を読み出す。そして、表様式・枠配置選択部は、当該表様式情報の中から当該抽出タイトルに対応する表様式を選択表様式として選択する。また、表様式・枠配置選択部は、当該選択表様式に対応する枠配置情報を表様式・枠配置格納部から読み出す。さらに、表様式・枠配置選択部は、特定項目格納部から特定項目を読み出す。そして、表様式・枠配置選択部は、当該特定項目に対応するデータ枠の表中での配置を当該

枠配置情報の中から特定枠配置として選択する。

【0030】また、表処理部は、先ず、表処理部は、表様式・特定枠配置選択部から特定枠配置を取得する。さらに、表領域・枠位置抽出部から枠位置情報を取得する。そして、表処理部は、当該特定枠配置に該当する位置の矩形枠を当該枠位置情報の中から特定矩形枠として選択する。また、表処理部は、当該特定矩形枠内から特定データ文字列領域を抽出する。続いて、表処理部は、当該特定データ文字列領域を文字認識部に転送する。そして、表処理部は、文字認識部から特定データを取得する。

【0031】また、文字認識部は、タイトル文字列領域抽出部からタイトル文字列領域を取得する。そして、文字認識部は、イメージ画像のうちのタイトル文字列領域について文字認識を行って、抽出タイトルを抽出する。抽出タイトルは、再びタイトル文字列領域抽出部へ転送される。また、文字認識部は、表処理部から特定データ文字列領域を取得する。そして、文字認識部は、イメージ画像のうちの特定データ文字列領域について文字認識を行って、特定データを抽出する。抽出された特定データは、再び表処理部へ転送される。

【0032】このように、この発明の第1の表処理装置によれば、上述の第1の表処理方法と同様に、表領域の全てについて文字認識を行わなくとも、特定データを選択的に抽出できる。従って、表領域の全てについて文字認識を行う場合に要する処理時間に比べて、表処理に要する処理時間の短縮を図ることができる。

【0033】また、この発明の第1の表処理装置において、好ましくは、枠配置情報は、各データ枠の配置を、表における行番号および列番号でそれぞれ表すのが良い。

【0034】また、この発明の第1の表処理装置の好適な構成例としては、表領域・枠位置抽出部は、イメージ画像を複数の方向にそれぞれ走査して、方向毎に、走査線毎の累積黒画素数を検出して周辺分布を作成する周辺分布作成部と、周辺分布に基づいて、イメージ画像から縦罫線および横罫線をそれぞれ抽出する罫線抽出部と、抽出された縦罫線および横罫線に囲まれた矩形枠をそれぞれ抽出し、抽出された矩形枠の位置を枠位置情報とする枠位置抽出部と、各矩形枠をもって構成される表領域を抽出する表領域抽出部とを具えてなることが望ましい。

【0035】さらに、この表処理装置において、枠位置情報は、矩形枠の位置と、表領域における当該矩形枠の行番号および列番号とを対応づけてあると良い。

【0036】また、この発明の第1の表処理装置の好適な構成例としては、表様式・特定枠配置選択部は、抽出タイトルと表様式情報に含まれるタイトルとを照合するタイトル照合部と、照合の結果、抽出タイトルと一致するタイトルに対応する表様式を、選択表様式として選択する表様式選択部と特定項目と、選択表様式に含まれる

項目とを照合する項目照合部と、照合の結果、当該特定項目に対応するデータ枠の表中での配置を枠配置情報の中から特定枠配置として選択する特定枠配置選択部とを具えてなることが望ましい。

【0037】尚、この発明の第1の表処理装置は、中央演算装置(CPU)、記憶手段および入出力装置を具えた、いわゆるコンピュータ装置のハードウェアを利用して構成されている。

【0038】(第2の表処理方法)また、この発明の第2の表処理方法によれば、表の特定項目に対応するデータ(以下、「特定データ」とも称する。)を当該表のイメージ画像から抽出するにあたり、まず、表様式格納部に、表のタイトルと表様式とを対応づけて表様式情報として予め格納しておく。さらに、この表様式格納部に、表様式毎に、当該表様式に属する項目の記載された項目枠と当該項目に対応するデータが記載されたデータ枠との配置関係を枠配置関係情報として予め格納しておく。

【0039】そして、第2の表処理方法では、(a)イメージ画像から、表領域を抽出すると共に、該表領域を構成する各矩形枠の位置を枠位置情報として抽出する。次に、(b)イメージ画像のうちの非表領域から表のタイトルを抽出タイトルとして抽出する。さらに、(c)表様式格納部から表様式情報を読み出して、当該表様式情報の中から抽出タイトルに対応する表様式を選択表様式として選択し、かつ、表様式格納部から当該選択表様式に対応する枠配置関係情報を選択枠配置関係情報として読み出す。続いて、(d)(a)の処理で抽出された枠位置情報の示す矩形枠の中から項目枠を選択する。そして、(e)項目枠の各々に記載された項目をそれぞれ抽出項目として抽出する。また、(f)抽出項目として特定項目が記載された項目枠を項目枠の中から特定項目枠として選択する。そして、(g)特定項目枠の位置に対して選択枠配置関係情報の示す位置にある矩形枠をイメージ画像から特定矩形枠として抽出する。最後に、(h)特定矩形枠内から特定データを抽出する。

【0040】そして、第2の表処理方法によれば、表領域の中の項目枠についてのみ文字認識を行うことにより、特定データを選択的に抽出できる。その結果、この表処理方法によれば、表領域の全てについて文字認識を行う場合に要する処理時間に比べて、表処理に要する処理時間の短縮を図ることができる。次に、特定データを選択的に抽出できる理由について説明する。

【0041】第2の表処理方法においては、タイトルと表様式とを対応づけて表様式情報として予め登録しておく。このように登録しておけば、イメージ画像から抽出したタイトル(抽出タイトル)を用いて、表様式を選択することができる。さらに、この処理方法では、表様式毎に、小見出し的な役割を持つ項目枠とデータ枠との配置関係を枠配置関係情報として予め登録しておく。このように登録しておけば、表様式を選択することにより、

その選択表様式における項目枠とデータ枠との配置関係(例えば、その選択表様式においては、項目枠の右側に隣接してデータ枠が配置されているという配置関係)を定めることができる。

【0042】そして、この表処理方法においては、表領域の矩形枠の中から項目枠を選択する。項目枠の選択にあたっては、後述する矩形枠の面積に着目する方法の他、任意好適な方法を用いると良い。続いて、選択された各項目枠について文字認識を行って、各項目枠からそれぞれ項目を抽出する。そして、各項目枠の中、特定項目が記載された特定項目枠を選択する。特定項目は、例えば予め指定しておくとも良い。

【0043】従って、この表処理方法においては、選択された特定項目枠の位置に対して、枠配置関係情報の示す配置関係に位置するデータ枠を特定矩形枠として選択することができる。すなわち、表領域の矩形枠のうち、項目枠についてのみ文字認識を行うことにより、特定データを選択的に抽出できる。その結果、表領域の全てについて文字認識を行わなくとも、特定データを選択的に抽出することができる。このため、表領域の全てについて文字認識を行ってから特定データを抽出する場合に要する処理時間に比べて、この処理方法では、特定データの抽出に要する処理時間の短縮を図ることができる。すなわち、表処理に要する時間の短縮を図ることができる。

【0044】また、第2の表処理方法によれば、枠配置関係情報として、項目枠とデータ枠との相対的な配置関係を登録している。すなわち、特定矩形枠の表中での配置は、1箇所に限定されない。このため、1つの表様式での特定矩形枠の配置は、枠配置関係情報の示す相対的な配置関係を満たす範囲内で任意の配置とすることができる。従って、第2の表処理方法によれば、特定矩形枠の配置を1箇所に限定した場合に処理対象となる表の種類の数よりも、より多くの種類の表を処理対象とすることができる。

【0045】次に、第2の表処理方法についてより詳細に説明する。

【0046】第2の表処理方法では、表処理の実行に先立ち、前述のように、タイトルと表様式とを対応づけて表様式情報として登録しておく。この表様式情報は、記憶部である表様式格納部に読み出し自在に格納しておく。また、各表様式毎に、当該表様式に属する項目の記載された項目枠と当該項目に対応するデータが記載されたデータ枠との配置関係を枠配置関係情報として登録しておく。この枠配置関係情報も、表様式情報と共に、表様式格納部に読み出し自在に格納しておく。

【0047】そして、この処理方法では、(a)まず、イメージ画像から表領域を抽出する。この際、表を構成する矩形枠の位置も枠位置情報として抽出する。

【0048】次に、この処理方法では、(b)そのイメ

ージ画像の中の非表領域、すなわち、表領域の外側の領域から表のタイトルを抽出する。

【0049】次に、この処理方法では、(c)抽出されたタイトル(抽出タイトル)を用いて、表様式を決定する。表様式の決定にあたっては、まず、表様式・枠配置格納部から表様式情報を読み出す。次に、抽出タイトルと、表様式情報に含まれるタイトルとを順次に照合する。照合の結果、抽出タイトルと一致したタイトルを選択する。そして、表様式情報から、この選択タイトルに対応づけられている表様式を選択表様式として選択する。さらに、当該選択表様式に対応する枠配置関係情報を表様式格納部から選択枠配置関係情報として読み出す。

【0050】また、この表処理方法では、(d)枠位置情報の示す矩形枠の中から項目枠を選択する。項目枠の選択にあたっては、任意好適な方法を用いると良い。例えば、後述のように、項目枠の面積がデータ枠の面積よりも小さい傾向に着目して項目枠を抽出しても良い。

【0051】そして、(e)項目枠の各々に記載された項目をそれぞれ抽出項目として抽出する。

【0052】また、(f)特定項目が抽出項目として記載された項目枠を項目枠の中から特定項目枠として選択する。特定項目枠の選択にあたっては、まず、特定項目と、抽出項目とを照合する。特定項目は、例えば特定項目格納部を設けて、そこに格納しておくとも良い。そして、照合の結果、特定項目と一致する項目が記載された項目枠を特定項目枠として、項目枠の中から選択する。

【0053】そして、(g)特定項目枠の位置に対して選択枠配置関係情報の示す位置にある矩形枠をイメージ画像から特定矩形枠として抽出する。

【0054】最後に、(h)特定矩形枠内から特定データを抽出する。

【0055】また、第2の表処理方法の実施にあたり、(a)の処理では、イメージ画像を複数の方向にそれぞれ走査して、方向毎に、走査線毎の累積黒画素数を検出して周辺分布を作成する処理と、周辺分布に基づいて、イメージ画像から縦罫線および横罫線をそれぞれ抽出する処理と、抽出された縦罫線および横罫線に囲まれた矩形枠をそれぞれ抽出し、抽出された矩形枠の位置を枠位置情報とする処理と、各矩形枠をもって構成される表領域を抽出する処理とを含むことが望ましい。

【0056】また、第2の表処理方法の実施にあたり、(b)の処理は、非表領域からタイトル文字列領域を抽出する処理と、タイトル文字列領域について文字認識処理を行って、抽出タイトルを抽出する処理とを含むことが望ましい。

【0057】また、第2の表処理方法において、好ましくは、(d)の処理は、矩形枠の面積を求める処理と、矩形枠のうち、一定面積以下の面積を有する矩形枠を項目枠として選択する処理とを含むことが望ましい。

【0058】これは、表領域の矩形枠のうち、項目枠は、データ枠よりも一般に面積が小さい傾向があることに着目した選択方法である。

【0059】また、第2の表処理方法において、好ましくは、(e)の処理は、項目枠から項目文字列領域を抽出する処理と、項目文字列領域について文字認識を行って、項目を抽出する処理とを含むことが望ましい。

【0060】また、第2の表処理方法において、好ましくは、(h)の処理は、特定矩形枠内から特定データ文字列領域を抽出する処理と、特定データ文字列領域について文字認識処理を行って、特定データを抽出する処理とを含むことが望ましい。

【0061】(第2の表処理装置)また、この発明の第2の表処理装置によれば、表の特定項目に対応するデータ(以下、「特定データ」とも称する。)を当該表のイメージ画像から抽出する表処理装置において、表様式格納部と、特定項目格納部と、表領域・枠位置抽出部と、タイトル文字列領域抽出部と、表様式選択部と、表処理部と、文字認識部とを具えている。

【0062】そして、表様式格納部は、表のタイトルと表様式とを対応づけて表様式情報として格納しておくと共に、表様式毎に、当該表様式に属する項目の記載された項目枠と当該項目に対応するデータが記載されたデータ枠との配置関係を枠配置関係情報として格納しておくための構成としてある。

【0063】また、特定項目格納部は、特定項目を格納しておくための構成としてある。

【0064】さらに、表領域・枠位置抽出部は、イメージ画像から、表領域を抽出すると共に、該表領域を構成する各矩形枠の位置を枠位置情報として抽出する構成としてある。

【0065】また、タイトル文字列領域抽出部は、イメージ画像のうちの非表領域からタイトル文字列領域を抽出し、当該タイトル文字列領域を文字認識部に転送し、文字認識部から抽出タイトルを取得する構成としてある。

【0066】また、表様式選択部は、まず、タイトル文字列領域抽出部から抽出タイトルを取得する。さらに、表様式選択部は、表様式格納部から表様式情報を読み出す。そして、表様式選択部は、当該表様式情報の中から当該抽出タイトルに対応する表様式を選択表様式として選択する。続いて、表様式選択部は、表様式格納部から当該選択表様式に対応する枠配置関係情報を選択枠配置関係情報として読み出す。

【0067】そして、表処理部は、まず、表領域・枠位置抽出部から枠位置情報を取得する。そして、表処理部は、当該枠位置情報の示す矩形枠の中から項目枠を選択する。続いて、表処理部は、項目枠の各々から項目文字列領域を抽出する。さらに、表処理部は、当該項目文字列領域を文字認識部に転送する。また、表処理部は、文

字認識部から抽出項目を取得する。続いて、表処理部は、特定項目格納部から特定項目を読み出す。そして、表処理部は、項目枠の中から当該特定項目が抽出項目として記載された項目枠を特定項目枠として選択する。さらに、表処理部は、表様式選択部から選択枠配置関係情報を取得する。そして、表処理部は、特定項目枠の位置に対して当該選択枠配置関係情報の示す位置にある矩形枠をイメージ画像から特定矩形枠として抽出する。また、表処理部は、特定矩形枠内から特定データ文字列領域を抽出する。続いて、表処理部は、当該特定データ文字列領域を文字認識部に転送する。そして、表処理部は、文字認識部から特定データを取得する。

【0068】さらに、文字認識部は、イメージ画像のうちのタイトル文字列領域について文字認識を行って抽出タイトルを抽出し、イメージ画像のうちの項目文字列領域について文字認識を行って項目を抽出し、かつ、イメージ画像のうちの特定データ文字列領域について字認識を行って特定データを抽出する構成としてある。

【0069】このように、この発明の第2の表処理装置によれば、上述の第2の表処理方法と同様に、表領域の中の項目枠についてのみ文字認識を行うことにより、特定データを選択的に抽出できる。その結果、この表処理方法によれば、表領域の全てについて文字認識を行う場合に要する処理時間に比べて、表処理に要する処理時間の短縮を図ることができる。

【0070】さらに、この発明の第2の表処理装置によれば、枠配置関係情報として、項目枠とデータ枠との相対的な配置関係を登録している。すなわち、特定矩形枠の表中での配置は、1箇所に限定されない。このため、1つの表様式での特定矩形枠の配置は、枠配置関係情報の示す相対的な配置関係を満たす範囲内で任意の配置とすることができる。従って、第2の表処理方法によれば、特定矩形枠の配置を1箇所に限定した場合に処理対象となる表の種類よりも、より多くの種類の表を処理対象とすることができる。

【0071】また、この発明の第2の表処理装置の好適な構成例としては、表領域・枠位置抽出部は、イメージ画像を複数の方向にそれぞれ走査して、方向毎に、走査線毎の累積黒画素数を検出して周辺分布を作成する処理と、周辺分布に基づいて、イメージ画像から縦罫線および横罫線をそれぞれ抽出する処理と、抽出された縦罫線および横罫線に囲まれた矩形枠をそれぞれ抽出し、抽出された矩形枠の位置を枠位置情報とする処理と、各矩形枠でもって構成される表領域を抽出する処理とを具備することが望ましい。

【0072】また、この発明の第2の表処理装置において、好ましくは、表処理部は、表領域から項目枠を選択するにあたり、各矩形枠の面積をそれぞれ求め、矩形枠のうち、一定面積以下の面積を有する矩形枠を項目枠として選択する構成を有するのが良い。

【0073】尚、この発明の第2の表処理装置は、中央演算装置（CPU）、記憶手段および入出力装置を具えた、いわゆるコンピュータ装置のハードウェアを利用して構成されている。

【0074】

【発明の実施の形態】以下、図面を参照して、この発明の第1および第2の表処理方法および装置の例について説明する。尚、参照する図は、この発明の構成および処理の流れが理解できる程度に概略的に示してあるにすぎない。従って、この発明は図示例にのみ限定されるものではない。

【0075】1. 第1の実施の形態

第1の実施の形態では、この発明の第1の表処理方法および装置の一例について説明する。

【0076】＜表処理装置＞まず、図1を参照して、第1の表処理装置の構成について説明する。図1は、第1の実施の形態の表処理装置の構成の説明に供する機能ブロック図である。

【0077】この実施の形態によれば、表の特定項目に対応するデータ（以下、「特定データ」とも称する。）を当該表のイメージ画像から抽出する表処理装置10は、表様式・枠配置格納部12と、特定項目格納部14と、表領域・枠位置抽出部16と、タイトル文字列領域抽出部18と、表様式・枠配置選択部20と、表処理部22と、文字認識部24とを具えている。

【0078】そして、この表処理装置10では、不図示の制御部によって、信号やデータの受け渡しの動作およびそのタイミングが制御されている。

【0079】また、この表処理装置10は、中央演算装置（CPU）、メモリ装置および入出力装置を具えたコンピュータのハードウェアを利用して構成されている。

【0080】また、この表処理装置10に入力されるイメージ画像は、この実施の形態では、画像読取部26によって読み取られたものである。この画像読取部26では、文書や帳票などの情報媒体の表面を光学的に走査して、画素毎の光信号の強度を量子化された電気信号に変換し、この電気信号からなるイメージ画像を出力する。イメージ画像は、例えば、文字部を黒画素で表現し、文字背景部を白画素で表現した2値画像とすると良い。

【0081】そして、この実施の形態では、読み取られたイメージ画像は、イメージ画像記憶部28に読み出し自在に格納される。格納されたイメージ画像には仮想的にx-y座標系が設定される。従って、イメージ画像の各画素の位置は、x座標およびy座標により与えられる。

【0082】以下、この表処理装置10の各構成成分について、それぞれ説明する。

【0083】（表様式・枠配置格納部）表様式・枠配置格納部12には、表のタイトルと表様式とを対応づけて表様式情報として格納してある。さらに、この表様式・

枠配置格納部12には、表様式毎に、当該表様式に属する項目と、当該項目に対応するデータが記載されたデータ枠の表領域中での配置とを対応づけて枠配置情報として格納してある。また、この実施の形態では、枠配置情報において各データ枠の配置を、表領域における行番号および列番号でそれぞれ表してある。このように、データ枠の配置を行番号および列番号を用いて表せば、表領域における枠の配置を容易かつ正確に表すことができる。

【0084】ここで、図2に、表様式・枠配置格納部12に格納されている、表様式情報および枠配置情報の例を示す。図2に示すように、この実施の形態では、表様式情報として、「注文書」、「依頼書」および「納品書」というタイトルをそれぞれ「様式1」、「様式2」および「様式3」という表様式と1対1に対応づけて格納してある。

【0085】さらに、この表様式・枠配置格納部12には、「注文書」の表様式である「様式1」に対応する枠配置情報として、「氏名」および「品名」という項目をそれぞれ「第1行第2列」および「第2行第2列」という、データ枠の配置と1対1にそれぞれ対応づけて格納してある。

【0086】また、この表様式・枠配置格納部12には、「依頼書」の表様式である「様式2」に対応する枠配置情報として、「氏名」、「住所」および「金額」という項目をそれぞれ「第1行第2列」、「第2行第2列」および「第3行第2列」という、データ枠の配置と1対1にそれぞれ対応づけて格納してある。

【0087】また、この表様式・枠配置格納部12には、「納品書」の表様式である「様式3」に対応する枠配置情報として、「品名」および「金額」という項目をそれぞれ「第2行第1列」および「第2行第2列」という、データ枠の配置と1対1にそれぞれ対応づけて格納してある。

【0088】(特定項目格納部) 特定項目格納部14には、特定項目を格納しておく。この実施の形態では、特定項目として、例えば「金額」を格納しておく。尚、特定項目は、表処理の途中で、例えばキーボードなどの入力装置を用いて入力しても良い。

【0089】(表領域・枠位置抽出部) 表領域・枠位置抽出部16は、処理の実行にあたり、まず、イメージ画像記憶部28からイメージ画像を読み出す。そして、表領域・枠位置抽出部16は、読み出されたイメージ画像から、表領域を抽出する。さらに、表領域・枠位置抽出部16は、イメージ画像からこの表領域を構成する各矩形枠の位置を枠位置情報として抽出する。

【0090】ここで、図3を参照して、表領域・枠位置抽出部16についてより詳細に説明する。図3は、表領域・枠位置抽出部16の説明に供する機能ブロック図である。図3に示すように、表領域・枠位置抽出部16

は、周辺分布作成部30、罫線抽出部32、枠位置抽出部34および表領域抽出部36をもって構成してある。

【0091】この周辺分布作成部30は、イメージ画像を複数の方向にそれぞれ走査して、方向毎に、走査線毎の累積黒画素数を検出して周辺分布を作成する。この実施の形態では、イメージ画像記憶部28から読み出されたイメージ画像をx軸方向に走査して、各走査線毎に累積画素数を検出して、横方向周辺分布を作成する。次に、イメージ画像をy軸方向に走査して、各走査線毎に累積画素数を検出して、縦方向周辺分布を作成する。

【0092】また、罫線抽出部32は、周辺分布に基づいて、イメージ画像から縦罫線および横罫線をそれぞれ抽出する。この実施の形態では、横方向周辺分布の各走査線毎の累積黒画素数の変化から、イメージ画像上での横罫線の両端の位置(座標)を検出する。また、縦方向周辺分布の各走査線毎の累積黒画素数の変化から、イメージ画像上での縦罫線の両端の位置(座標)を検出する。そして、イメージ画像における各罫線の位置を、それぞれその両端の座標をもって表す。

【0093】また、枠位置抽出部34は、抽出された縦罫線および横罫線に囲まれた矩形枠をそれぞれ抽出する。そして、抽出された矩形枠の位置を枠位置情報とする。この実施の形態では、互いに隣り合った2本の縦罫線と互いに隣り合った2本横罫線との交点である4つ一組の交点の位置を、矩形枠の4頂点の位置として抽出する。そして、枠位置抽出部34は、イメージ画像における各矩形枠の位置を、それぞれその4頂点の座標をもって表した枠位置情報を作成する。

【0094】さらに、この実施の形態では、各矩形枠の位置から、当該矩形枠の表中の行番号および列番号を求める。そして、枠位置情報において、各矩形枠の位置と、当該矩形枠の行番号および列番号とを対応づけておく。

【0095】また、表領域抽出部36は、各矩形枠をもって構成される表領域を抽出する。表領域は、例えば、縦罫線のうちの最外側の2本の縦罫線と、横罫線のうちの最外側の2本の横罫線とによって囲まれた領域を表領域として抽出すると良い。もしくは、各矩形枠の集合を求めて、その集合のイメージ画像における輪郭を表領域として抽出すると良い。また、イメージ画像における表領域の位置は、表領域の4つ頂点の座標をもって表すと良い。

【0096】(タイトル文字列領域抽出部) タイトル文字列領域抽出部18は、イメージ画像のうちの非表領域からタイトル文字列領域を抽出する。タイトル文字列領域の抽出にあたっては、従来公知の技術を用いることができる。例えば、表領域・枠位置抽出部16で抽出した表領域の外側周囲のイメージ画像部分の周辺分布を作成して、黒画素の連結成分に基づいてタイトル文字列領域を抽出する。

【0097】そして、タイトル文字列領域抽出部18は、抽出した当該タイトル文字列領域を文字認識部24に転送する。この実施の形態では、タイトル文字列領域の外接座標を求め、この外接座標を文字認識部24に転送する。

【0098】さらに、タイトル文字列領域抽出部18は、文字認識部24から、文字認識結果である抽出タイトルを取得する。この実施の形態では、文字認識結果として文字コードを取得する。

【0099】(表様式・枠配置選択部)また、表様式・枠配置選択部20は、まず、タイトル文字列領域抽出部18から抽出タイトルを取得する。さらに、表様式・枠配置選択部20は、表様式・枠配置格納部12から表様式情報を読み出す。そして、表様式・枠配置選択部20は、当該表様式情報の中から当該抽出タイトルに対応する表様式を選択表様式として選択する。また、表様式・枠配置選択部20は、当該選択表様式に対応する枠配置情報を表様式・枠配置格納部12から読み出す。さらに、表様式・枠配置選択部20は、特定項目格納部14から特定項目を読み出す。そして、表様式・枠配置選択部20は、当該特定項目に対応するデータ枠の表中での配置を枠配置情報の中から特定枠配置として選択する。

【0100】ここで、図4を参照して、表様式・枠配置選択部20についてより詳細に説明する。図4は、表様式・枠配置選択部20の説明に供する機能ブロック図である。図4に示すように、表様式・枠配置選択部20は、タイトル照合部38、表様式選択部40、項目照合部42および特定枠配置選択部44を具えている。

【0101】このタイトル照合部38は、文字認識部24で得られた抽出タイトルの文字コードを、タイトル文字列領域抽出部18を介して取得する。また、このタイトル照合部38は、表様式・枠配置格納部12から表様式情報を読み出す。この表様式情報においては、前述したように、タイトルと表様式とが対応づけられている。そして、タイトル照合部38は、抽出タイトルと表様式情報に含まれるタイトルとを照合する。

【0102】また、表様式選択部40は、照合の結果、抽出タイトルと一致するタイトルに対応する表様式を、選択表様式として選択する。

【0103】また、項目照合部42は、特定項目格納部14から特定項目を読み出す。また、表様式選択部40で選択された選択表様式には、前述したように、1つもしくは複数の項目が含まれている。そこで、項目照合部42は、読み出された特定項目と、選択表様式に含まれる項目とを照合する。

【0104】また、特定枠配置選択部44は、照合の結果、特定項目と一致する項目に対応する特定矩形枠の特定枠配置を枠配置情報の中から選択する。この特定枠配置は、前述したように、表領域中の行番号および列番号で表されている。

【0105】(表処理部)また、表処理部22は、先ず、表様式・枠配置選択部20から特定枠配置を取得する。さらに、表処理部22は、表領域・枠位置抽出部16から枠位置情報を取得する。

【0106】そして、表処理部22は、当該特定枠配置に該当する位置の矩形枠を枠位置情報の中から特定矩形枠として選択する。特定矩形枠の選択にあたっては、先ず、特定枠配置の行番号および列番号を取得する。そして、枠位置情報には、前述したように、各矩形枠の位置(座標)とその行番号および列番号とがそれぞれ対応づけられている。そこで、この実施の形態では、特定枠配置の行および列番号と、枠位置情報の矩形枠の行および列番号とを照合する。そして、照合の結果、行および列番号の一致した矩形枠の位置を、イメージ画像における特定矩形枠の位置とする。特定矩形枠の位置は、その矩形枠の4頂点の位置の座標で与えられる。

【0107】また、表処理部22は、当該特定矩形枠内から特定データ文字列領域を抽出する。特定データ文字列領域の抽出にあたっては、従来公知の技術を用いることができる。例えば、表領域・枠位置抽出部16で抽出した矩形枠の内部のイメージ画像部分の周辺分布を作成して、黒画素の連結成分に基づいて特定データ文字列領域を抽出する。

【0108】続いて、表処理部22は、当該特定データ文字列領域を文字認識部24に転送する。この実施の形態では、特定データ文字列領域の外接座標を求め、この外接座標を文字認識部24に転送する。

【0109】そして、表処理部22は、文字認識部24から特定データを取得する。この実施の形態では、文字認識結果として文字コードを取得する。

【0110】そして、表処理部22は、特定データを表処理結果として、表処理装置10の外部へ出力する。出力にあたっては、例えばディスプレイなどの出力装置を用いると良い。

【0111】(文字認識部)また、文字認識部24は、文字認識にあたり、文字認識部24の内部メモリ(不図示)に記憶されている辞書を参照する。そして、文字認識部24は、辞書に格納されている標準文字のうち、認識対象と一致する標準文字の文字コードを認識結果として出力する。文字コードには、例えば、JISコードを用いると良い。

【0112】例えば、抽出タイトルを抽出する場合は、文字認識部24は、前述したように、タイトル文字列領域抽出部18からタイトル文字列領域を取得する。そして、文字認識部24は、イメージ画像のうちのタイトル文字列領域について文字認識を行って、抽出タイトルを文字コードとして抽出する。抽出タイトルは、前述したように、再びタイトル文字列領域抽出部18へ転送される。

【0113】また、特定データを抽出する場合は、文字

認識部24は、表処理部22から特定データ文字列領域を取得する。そして、文字認識部24は、イメージ画像のうちの特定データ文字列領域について文字認識を行って、特定データを文字コードとして抽出する。抽出された特定データは、前述したように、再び表処理部22へ転送される。

【0114】<表処理方法>次に、図5を参照して、第1の実施の形態の表処理方法、即ち、図1に示す表処理装置の動作の一例について説明する。図5は、第1の実施の形態の表処理方法の説明に供するフローチャートである。また、この実施の形態では、図12に示す「依頼書」の表を処理対象として、この「依頼書」のうちの特定項目「金額」のデータである「15,000」円を特定データとして抽出する例について説明する。尚、各ステップの処理の開始のタイミングは、例えば、不図示の制御部によってコントロールすると良い。

【0115】第1の実施の形態の表処理方法によれば、表の特定項目に対応するデータ（以下、「特定データ」とも称する。）を当該表のイメージ画像から抽出するにあたり、まず、表様式・枠配置格納部12に、表のタイトルと表様式とを対応づけて表様式情報として予め格納しておく。さらに、この表様式・枠配置格納部12に、表様式毎に、当該表様式に属する項目と、当該項目に対応するデータが記載されたデータ枠の表中での配置とを対応づけて枠配置情報として予め格納しておく。

【0116】（ステップ1）そして、第1の実施の形態の表処理方法では、まず、表領域・枠位置抽出部16によって、イメージ画像から、表領域を抽出すると共に、この表領域を構成する各矩形枠の位置を枠位置情報として抽出する（図5のS1）。

【0117】ここで、図6を参照して、表領域および矩形枠の位置を抽出する処理ステップ（S1）について説明する。図6は、表領域および矩形枠の位置を抽出する処理の説明に供するフローチャートである。

【0118】この実施の形態では、表領域および矩形枠の位置を抽出するにあたり、まず、表領域・枠位置抽出部16の周辺分布作成部30において、イメージ画像を複数の方向にそれぞれ走査して、方向毎に、走査線毎の累積黒画素数を検出して周辺分布を作成する（図6のS7）。この実施の形態では、イメージ画像記憶部28から読み出されたイメージ画像をx軸方向に走査して、各走査線毎に累積画素数を検出して、横方向周辺分布を作成する。次に、イメージ画像をy軸方向に走査して、各走査線毎に累積画素数を検出して、縦方向周辺分布を作成する。

【0119】次に、表領域・枠位置抽出部16の罫線抽出部32が、周辺分布に基づいて、イメージ画像から縦罫線および横罫線をそれぞれ抽出する（図6のS8）。この実施の形態では、横方向周辺分布の各走査線毎の累積黒画素数の変化から、イメージ画像上での横罫線の両

端の位置（座標）を検出する。また、縦方向周辺分布の各走査線毎の累積黒画素数の変化から、イメージ画像上での縦罫線の両端の位置（座標）を検出する。そして、イメージ画像における各罫線の位置を、それぞれその両端の座標をもって表す。

【0120】次に、表領域・枠位置抽出部16の枠位置抽出部34が、抽出された縦罫線および横罫線に囲まれた矩形枠の位置をそれぞれ抽出する（図6のS9）。そして、抽出された矩形枠の位置を枠位置情報とする。この実施の形態では、互いに隣り合った2本の縦罫線と互いに隣り合った2本横罫線との交点である4つ一組の交点の位置を、矩形枠の4頂点の位置として抽出する。そして、枠位置抽出部34は、イメージ画像における各矩形枠の位置を、それぞれその4頂点の座標をもって表した枠位置情報を作成する。

【0121】尚、この実施の形態では、各矩形枠の位置から、当該矩形枠の表中の行番号および列番号を求める。そして、枠位置情報において、各矩形枠の位置と、当該矩形枠の行番号および列番号とを対応づけておく。

【0122】次に、表領域・枠位置抽出部16の表領域抽出部36が、各矩形枠でもって構成される表領域58（図12参照）を抽出する（図6のS10）。表領域は、例えば、縦罫線のうちの最外側の2本の縦罫線と、横罫線のうちの最外側の2本の横罫線とによって囲まれた領域を表領域として抽出すると良い。もしくは、各矩形枠の集合を求めて、その集合のイメージ画像における輪郭を表領域として抽出すると良い。また、イメージ画像における表領域の位置は、表領域の4つ頂点の座標をもって表すと良い。

【0123】このようにして、表領域・枠位置抽出部16によって、表領域58および矩形枠の位置を抽出する（図5のS1）。

【0124】（ステップ2）次に、タイトル文字列領域抽出部18および文字認識部24によって、イメージ画像のうちの非表領域から、表のタイトルを抽出タイトルとして抽出する（図5のS2）。

【0125】ここで、図7を参照して、抽出タイトルを抽出する処理ステップ（S2）について説明する。図7は、抽出タイトルの抽出処理の説明に供するフローチャートである。

【0126】この実施の形態では、抽出タイトルを抽出するにあたり、まず、タイトル文字列領域抽出部18において、表領域・枠位置抽出部16で抽出した表領域の外側周囲のイメージ画像部分の周辺分布を作成して、黒画素の連結成分に基づいてタイトル文字列領域を抽出する（図7のS11）。そして、タイトル文字列領域抽出部18は、抽出したタイトル文字列領域に外接する外接矩形枠の座標を文字認識部24へ転送する。尚、外接矩形枠の座標は、外接矩形枠の4頂点のx-y座標として表すと良い。

【0127】次に、文字認識部24は、外接矩形枠の座標に対応するイメージ画像について文字認識を行う(図7のS12)。文字認識部24は、文字認識にあたり、文字認識部24の内部メモリ(不図示)に記憶されている辞書を参照する。そして、文字認識部24は、辞書に格納されている標準文字のうち、認識対象と一致する標準文字の文字コードを認識結果として出力する。そして、文字認識部24は、この文字コードをタイトル文字列領域抽出部18へ転送する。この実施の形態では、文字認識の結果、「依頼書」という抽出タイトルを抽出したとする。

【0128】タイトル文字列領域抽出部18は、この「依頼書」の文字コードを抽出タイトルとして表様式・枠配置選択部20へ転送する。

【0129】このようにして、タイトル文字列領域抽出部18および文字認識部24によって、抽出タイトルを抽出する。

【0130】(ステップ3)次に、表様式・枠配置選択部20によって、表様式・枠配置格納部12から表様式情報を読み出して、当該表様式情報の中から抽出タイトルに対応する表様式を選択表様式として選択する(図5のS3)。

【0131】ここで、図8を参照して、選択表様式を選択する処理ステップ(S3)について説明する。図8は、選択表様式の選択処理の説明に供するフローチャートである。

【0132】この実施の形態では、選択表様式を選択するにあたり、まず、表様式・枠配置選択部20のタイトル照合部38が、タイトル文字列領域抽出部18から抽出タイトルを取得する(図8のS13)。ここでは、抽出タイトルとして、「依頼書」の文字コードを取得する。尚、ステップ3の処理は、例えば、タイトル文字列領域抽出部18から表様式・枠配置選択部20へ文字コードが転送されたことをもって開始させても良い。

【0133】次に、タイトル照合部38は、表様式・枠配置格納部12から、表様式情報を1つずつ順次を読み出す(図8のS14)。この実施の形態では、前述したように、表様式・枠配置格納部12に図2に示した表様式情報を格納している。この実施の形態では、タイトル照合部38は、まず、1番目の表様式情報として「注文書-様式1」を読み出す。

【0134】次に、タイトル照合部38は、抽出タイトルと表様式情報とを照合する(図8のS15)。

【0135】ここでは、まず、抽出タイトルである「依頼書」の文字コードと、1回目に読み出された表様式情報のタイトルである「注文書」の文字コードとを照合する。この場合、1回目の照合結果は一致しない。

【0136】照合結果が一致しない場合(S15のステップの「no」の場合)、タイトル照合部38は、次の表様式情報へアクセスして(図8のS16)、2番目の

表様式情報として「依頼書-様式2」を読み出す(図8のS14)。そして、1回目と同様に、抽出タイトルである「依頼書」の文字コードと、2回目に読み出された表様式のタイトルである「依頼書」の文字コードとを照合する。この場合の照合結果は一致する。

【0137】照合結果が一致した場合には、タイトル照合部38は、照合結果が一致した表様式情報を表様式選択部40へ転送する。

【0138】次に、表様式・枠配置選択部20の表様式選択部40は、表様式を決定する(図8のS17)。この実施の形態では、タイトル照合部38は、タイトル照合部38から取得した表様式情報の示す様式「様式2」を確認し、処理対象の表領域の表様式を「様式2」と決定する。

【0139】この実施の形態では、上述した、表様式選択処理の照合処理(S15)は、全ての表様式情報について同様に行われる。そして、選択表様式が決定される度に、制御部(不図示)が、表様式・枠配置格納部12に格納されている全表様式数を参照して、全ての表様式について照合が行われたか否かを判定する(図8のS18)。そして、全ての表様式について処理が行われていないと判定した場合(S18の「no」)には、次の表様式情報にアクセスして(図8のS19)、タイトル照合部38にその表様式情報を読み出させる(図8のS14)。そして、再び表様式情報と抽出タイトルとの照合処理(S15)を行う。

【0140】尚、この発明では、全ての表処理情報についての照合は必ずしも必要ではなく、例えば、選択表様式を決定したステップ(S17)の後直ちに表様式選択処理を終了しても良い。

【0141】また、制御部が、全ての表様式について処理が行われたと判定した場合(S18の「yes」)には、表様式選択処理を終了する。

【0142】(ステップ4)次に、表様式・枠配置選択部20によって、特定データの記載されている矩形枠の配置としての特定枠配置を選択する(図5のS4)。

【0143】ここで、図9を参照して、特定枠配置を選択する処理ステップ(S4)について説明する。図9は、特定枠配置を選択する処理の説明に供するフローチャートである。

【0144】この実施の形態では、特定枠配置を選択するにあたり、まず、表様式・枠配置選択部20の項目照合部42が、選択表様式に対応する枠配置情報を表様式・枠配置格納部12から取得する(図9のS20)。ここでは、「様式2」に対応する枠配置情報(図2参照)を読み出す。

【0145】次に、項目照合部42は、特定項目格納部14から特定項目を読み出す(図9のS21)。ここでは、処理対象の表から得ようとする情報の項目である特定項目として、特定項目格納部14に格納されていた

「金額」を読み出す。

【0146】次に、項目照合部42は、特定項目「金額」と枠配置情報の各項目とを1つずつ順次に照合する(図9のS22)。ここでは、特定項目「金額」と、「様式2」に対応する枠配置情報の1番目の項目「氏名」とを照合する。

【0147】そして、照合の結果、特定項目と枠配置情報の項目とが一致しない場合(S22の「no」の場合)には、枠配置情報の次の項目にアクセスして(図9のS23)、再び照合処理(S22)を行う。ここでは、1番目の項目「氏名」と特定項目「金額」とは一致しない。このため、次の2番目の項目「住所」について照合処理(S22)を行う。この場合、2番目の項目「住所」と特定項目「金額」とも一致しない。このため、3番目の項目「金額」について照合処理(S23、S22)を行う。

【0148】そして、照合の結果、特定項目と枠配置情報の項目とが一致した場合(S22の「yes」の場合)には、項目照合部42は、枠配置情報の項目のうち特定項目と一致した項目を、表様式・枠配置選択部20の特定枠配置選択部44へ転送する。ここでは、3番目の項目「金額」と特定項目「金額」とが一致する。従って、項目照合部42は、3番目の項目「金額」を特定枠配置選択部44へ転送する。

【0149】次に、特定枠配置選択部44によって、当該特定項目に対応するデータ枠の表中での配置を枠配置情報の中から特定枠配置として選択する(図9のS24)。この実施の形態では、特定枠配置選択部44は、「様式2」に対応する枠配置情報の「金額」に対応するデータ枠の配置を特定枠配置として選択する。ここでは、データ枠の配置は、表領域中の行番号および列番号で表されている。具体的には、特定枠配置は「第3行第2列」となる(図2参照)。

【0150】この実施の形態では、上述した、特定項目と枠配置情報の項目との照合処理(S22)は、「様式2」に対応する枠配置情報の全ての項目について同様に行われる。そして、特定枠配置が決定される度に、制御部(不図示)が、表様式・枠配置格納部12に格納されている、「様式2」に対応する枠配置情報に含まれる全項目数を参照して、全ての項目について照合が行われたか否かを判定する(図9のS25)。そして、全ての項目について処理が行われていないと判定した場合(S25の「no」の場合)には、次の項目にアクセスして(図9のS26)、項目照合部42にその項目と特定項目との照合処理(S22)を行わせる。

【0151】尚、この発明では、全ての項目についての照合は必ずしも必要ではなく、例えば、特定枠配置を決定したステップ(S24)の後直ちに特定枠配置選択処理を終了しても良い。

【0152】また、制御部が、全ての項目について処理

が行われたと判定した場合(S25の「yes」の場合)には、特定枠配置選択処理を終了する。

【0153】そして、表様式・枠配置選択部20は、特定枠配置を表す「第3行第2列」を表処理部22へ転送する。

【0154】(ステップ5)ところで、この特定枠配置は、特定のデータ枠の表様式における配置を示しているにすぎない。一方、イメージ画像における各矩形枠の位置は、処理ステップS1において枠位置情報として求められている。

【0155】そこで、次に、表処理部22によって、枠位置情報の中から、イメージ画像において特定枠配置に該当する位置の矩形枠を特定矩形枠として選択する(図5のS5)。

【0156】ここで、図10を参照して、特定矩形枠を選択する処理ステップ(S5)について説明する。図10は、特定矩形枠を選択する処理の説明に供するフローチャートである。

【0157】この実施の形態では、特定矩形枠を選択するにあたり、表様式・特定枠配置選択部20から特定枠配置を取得した表処理部22は、先ず、表領域・枠位置抽出部16から枠位置情報を取得する(図10のS27)。枠位置情報には、前述したように、各矩形枠の位置(座標)とその行番号および列番号とがそれぞれ対応づけられている。また、特定枠配置も行番号および列番号で与えられる。

【0158】次に、表処理部22は、当該特定枠配置に該当する位置の矩形枠を枠位置情報の中から特定矩形枠として選択する(図10のS28およびS30)。

【0159】そのために、この実施の形態では、先ず、特定枠配置の行および列番号(行列番号)と、枠位置情報の矩形枠の行および列番号(行列番号)とを1つずつ順に照合する(図10のS28)。

【0160】そして、照合の結果、特定枠配置の行列番号と枠位置情報の行列番号とが一致しない場合(S28の「no」の場合)には、枠位置情報の次の項目にアクセスして(図10のS29)、再び照合処理(S28)を行う。

【0161】そして、照合の結果、特定枠配置の行列番号と枠位置情報の行列番号とが一致した場合(S28の「yes」の場合)には、表処理部22は、枠位置情報のうち行列番号が一致した矩形枠の位置を、図12に示すイメージ画像における特定矩形枠60の位置として決定する(図10のS30)。特定矩形枠の位置は、その矩形枠の4頂点の位置の座標で与えられる。

【0162】この実施の形態では、枠位置情報のうちの行列番号が「第3行第2列」の矩形枠が特定矩形枠として選択される。

【0163】この実施の形態では、上述した、特定枠配置の行列番号と枠位置情報の行列番号との照合処理(S

28)は、全ての矩形枠について同様に行われる。そして、特定矩形枠が決定される度に、制御部(不図示)が、枠位置情報を参照して、全ての矩形枠について照合が行われたか否かを判定する(図9のS31)。そして、全ての矩形枠について処理が行われていないと判定した場合(S31の「no」の場合)には、次の矩形項目にアクセスして(図10のS32)、次の枠位置情報を取得する(S27)。そして、再び表処理部22に次の矩形枠の行列番号と特定枠配置の行列番号との照合処理(S28)を行わせる。

【0164】尚、この発明では、全ての矩形枠の行列番号についての照合は必ずしも必要ではなく、例えば、特定矩形枠を決定したステップ(S30)の後直ちに特定矩形枠選択処理を終了しても良い。

【0165】また、制御部(不図示)が、全ての矩形枠について処理が行われたと判定した場合(S31の「yes」の場合)には、特定矩形枠選択処理を終了する。

【0166】(ステップ6)次に、表処理部22によって、特定矩形枠内から特定データを抽出する(図5のS6)。

【0167】ここで、図11を参照して、特定データを抽出する処理ステップ(S6)について説明する。図11は、特定データを抽出する処理の説明に供するフローチャートである。

【0168】この実施の形態では、特定データを抽出するにあたり、表処理部22は、先ず、「第3行第2列」の特定矩形枠60(図12参照)の領域から特定データ文字列領域を抽出する(図11のS33)。特定データ文字列領域の抽出にあたっては、従来公知の技術を用いることができる。例えば、表領域・枠位置抽出部16で抽出した矩形枠の内部のイメージ画像部分の周辺分布を作成して、黒画素の連結成分に基づいて特定データ文字列領域を抽出する。

【0169】続いて、表処理部22は、抽出したデータ文字列領域に外接する外接矩形枠の座標を文字認識部24へ転送する。尚、外接矩形枠の座標は、外接矩形枠の4頂点のx-y座標として表すと良い。

【0170】次に、文字認識部24は、外接矩形枠の座標に対応するイメージ画像について文字認識を行う(図11のS34)。文字認識部24は、文字認識にあたり、文字認識部24の内部メモリ(不図示)に記憶されている辞書を参照する。そして、文字認識部24は、辞書に格納されている標準文字のうち、認識対象と一致する標準文字の文字コードを認識結果として出力する。そして、文字認識部24は、認識結果としての文字コードを表処理部22へ転送する。この実施の形態では、文字認識の結果、「15,000」という金額の特定データを抽出したとする。

【0171】続いて、表処理部22は、この「15,000」の文字コードを特定データとして表処理装置10

の外部へ出力する。出力にあたっては、例えばディスプレイなどの出力装置を用いると良い。

【0172】このように、第1の実施の形態の表処理方法によれば、表領域の全てについて文字認識を行わなくとも、特定データを選択的に抽出できる。従って、表領域の全てについて文字認識を行う場合に要する処理時間に比べて、表処理に要する処理時間の短縮を図ることができる。

【0173】2. 第2の実施の形態

第2の実施の形態では、この発明の第2の表処理方法および装置の一例について説明する。

【0174】<表処理装置>先ず、図13を参照して、第2の実施の形態の表処理装置の構成について説明する。図13は、第2の実施の形態の表処理装置の構成の説明に供する機能ブロック図である。

【0175】この実施の形態によれば、表の特定項目に対応するデータ(以下、「特定データ」とも称する。)を当該表のイメージ画像から抽出する表処理装置46は、表様式格納部48と、特定項目格納部14と、表領域・枠位置抽出部16と、タイトル文字列領域抽出部18と、表様式選択部50と、表処理部54と、文字認識部52とを具えている。

【0176】そして、この表処理装置46では、不図示の制御部によって、信号やデータの受け渡しの動作およびそのタイミングが制御されている。

【0177】また、この表処理装置46は、中央演算装置(CPU)、メモリ装置および入出力装置を具えたコンピュータのハードウェアを利用して構成されている。

【0178】また、この表処理装置46に入力されるイメージ画像は、上述の第1の実施の形態の場合と同様に、画像入力部26によって読み取られた後、イメージ画像記憶部28に読み出し自在に記憶されたものである。

【0179】以下、この表処理装置46の各構成成分について、それぞれ説明する。

【0180】(表様式格納部)表様式格納部48には、表のタイトルと表様式とを対応づけて表様式情報として格納している。さらに、この表様式格納部48には、表様式毎に、当該表様式に属する項目の記載された項目枠と当該項目に対応するデータが記載されたデータ枠との配置関係を枠配置関係情報として格納している。

【0181】ここで、図14に、表様式格納部48に格納されている、表様式情報および枠配置関係情報の例を示す。図14に示すように、この実施の形態では、表様式情報として、「注文書」および「依頼書」というタイトルをそれぞれ「様式1」および「様式2」という表様式と1対1に対応づけて格納している。

【0182】さらに、この表様式格納部48には、表様式毎に、項目枠とデータ枠の配置関係を枠配置関係情報として格納している。図14に示すように、この実施の

形態では、「様式1」の場合は「項目枠の下側にデータ枠がある」という枠配置関係情報が格納されている。また、「様式2」の場合は、「項目枠の右側にデータ枠がある」という枠配置関係情報が格納されている。

【0183】このように、第2の実施の形態では、枠配置関係情報として、項目枠とデータ枠との相対的な配置関係を登録している。すなわち、特定矩形枠の表中での配置は、1箇所に限定されない。このため、1つの表様式での特定矩形枠の配置は、枠配置関係情報の示す相対的な配置関係を満たす範囲内で任意の配置とすることができる。従って、第2の実施の形態の表処理装置によれば、特定矩形枠の配置を1箇所に限定した場合に処理対象となる表の種類の数よりも、より多くの種類の表を処理対象とすることができる。

【0184】(特定項目格納部) 特定項目格納部14には、特定項目を格納しておく。この実施の形態では、特定項目として、例えば「金額」を格納しておく。尚、特定項目は、表処理の途中で、例えばキーボードなどの入力装置を用いて入力しても良い。

【0185】(表領域・枠位置抽出部) 表領域・枠位置抽出部16は、処理の実行にあたり、先ず、イメージ画像記憶部28からイメージ画像を読み出す。そして、表領域・枠位置抽出部16は、読み出されたイメージ画像から、表領域を抽出する。さらに、表領域・枠位置抽出部16は、イメージ画像からこの表領域を構成する各矩形枠の位置を枠位置情報として抽出する。尚、第2の実施の形態の表領域・枠位置抽出部16は、上述した第1の実施の形態の場合と同一の構成であるので、その詳細な説明を省略する。

【0186】(タイトル文字列領域抽出部) タイトル文字列領域抽出部18は、イメージ画像のうちの非表領域からタイトル文字列領域を抽出する。続いて、タイトル文字列領域抽出部18は、抽出した当該タイトル文字列領域を文字認識部24に転送する。さらに、タイトル文字列領域抽出部18は、文字認識部24から、文字認識結果である抽出タイトルを取得する。尚、第2の実施の形態のタイトル文字列領域抽出部18は、上述した第1の実施の形態の場合と同一の構成であるので、その詳細な説明を省略する。

【0187】(表様式選択部) また、表様式選択部50は、先ず、タイトル文字列領域抽出部18から抽出タイトルを取得する。さらに、表様式選択部50は、表様式格納部48から表様式情報を読み出す。そして、表様式選択部50は、当該表様式情報の中から当該抽出タイトルに対応する表様式を選択表様式として選択する。続いて、表様式選択部50は、表様式格納部48から該選択表様式に対応する枠配置関係情報を選択枠配置関係情報として読み出す。

【0188】(表処理部) そして、表処理部54は、先ず、表領域・枠位置抽出部16から枠位置情報を取得す

る。そして、表処理部54は、当該枠位置情報の示す矩形枠の中から項目枠を選択する。この実施の形態では、表領域から項目枠を選択するにあたり、表処理部54は、先ず、各矩形枠の面積をそれぞれ求める。次に、表処理部54は、矩形枠のうち、一定面積以下の面積を有する矩形枠を項目枠として選択する。

【0189】続いて、表処理部54は、項目枠の各々から項目文字列領域を抽出する。項目文字列領域の抽出にあたっては、従来公知の技術を用いることができる。この実施の形態では、例えば、表領域・枠位置抽出部16で抽出した矩形枠の内部のイメージ画像部分の周辺分布を作成して、この周辺分布における黒画素の連結成分に基づいて各項目文字列領域をそれぞれ抽出する。

【0190】さらに、表処理部54は、当該項目文字列領域を文字認識部52に転送する。この実施の形態では、各項目文字列領域の外接座標をそれぞれ求め、この外接座標をそれぞれ文字認識部52に転送する。

【0191】また、表処理部54は、文字認識部52から抽出項目を取得する。この実施の形態では、文字認識結果として各項目の文字コードをそれぞれ取得する。

【0192】続いて、表処理部54は、特定項目格納部14から特定項目を読み出す。

【0193】そして、表処理部54は、項目枠の中から当該特定項目が抽出項目として記載された項目枠を特定項目枠として選択する。

【0194】さらに、表処理部54は、表様式選択部50から選択枠配置関係情報を取得する。配置関係情報とは、上述したように、例えば「項目枠の右側にデータ枠がある」というものである。

【0195】そして、表処理部54は、特定項目枠の位置に対して当該選択枠配置関係情報の示す位置にある矩形枠をイメージ画像から特定矩形枠として抽出する。例えば、特定項目枠が表の3行1列目の矩形枠である場合、その矩形枠の右側の3行2列目の矩形枠が、特定矩形枠として抽出される。

【0196】また、表処理部54は、特定矩形枠内から特定データ文字列領域を抽出する。特定データ文字列領域の抽出にあたっては、従来公知の技術を用いることができる。この実施の形態では、例えば、表領域・枠位置抽出部16で抽出した矩形枠の内部のイメージ画像部分の周辺分布を作成して、この周辺分布における黒画素の連結成分に基づいて特定データ文字列領域を検出する。

【0197】続いて、表処理部54は、当該特定データ文字列領域を文字認識部52に転送する。この実施の形態では、特定データ文字列領域の外接座標を求め、この外接座標を文字認識部52に転送する。

【0198】そして、表処理部54は、文字認識部52から特定データを取得する。この実施の形態では、文字認識結果として特定データの文字コードを取得する。

【0199】このように、第2の実施の形態では、表領

域の中の項目枠についてのみ文字認識を行うことにより、特定データを選択的に抽出できる。その結果、この表処理装置によれば、表領域の全てについて文字認識を行う場合に要する処理時間に比べて、表処理に要する処理時間の短縮を図ることができる。

【0200】(文字認識部) また、文字認識部52は、文字認識にあたり、文字認識部52の内部メモリ(不図示)に記憶されている辞書を参照する。そして、文字認識部52は、辞書に格納されている標準文字のうち、認識対象と一致する標準文字の文字コードを認識結果として出力する。文字コードには、例えば、JISコードを用いると良い。

【0201】例えば、抽出タイトルを抽出する場合は、文字認識部52は、前述したように、タイトル文字列領域抽出部18からタイトル文字列領域を取得する。そして、文字認識部52は、イメージ画像のうちのタイトル文字列領域について文字認識を行って、抽出タイトルを文字コードとして抽出する。抽出タイトルは、前述したように、再びタイトル文字列領域抽出部18へ転送される。

【0202】また、項目を抽出する場合は、文字認識部52は、表処理部54から項目文字列領域を取得する。そして、文字認識部52は、イメージ画像のうちの項目文字列領域について文字認識を行って、項目を文字コードとして抽出する。抽出された項目は、前述したように、再び表処理部54へ転送される。

【0203】また、特定データを抽出する場合は、文字認識部52は、表処理部54から特定データ文字列領域を取得する。そして、文字認識部52は、イメージ画像のうちの特定データ文字列領域について文字認識を行って、特定データを文字コードとして抽出する。抽出された特定データは、前述したように、再び表処理部54へ転送される。

【0204】<表処理方法>次に、図15を参照して、第2の実施の形態の表処理方法、すなわち、図13に示す表処理装置の動作の一例について説明する。図15は、第2の実施の形態の表処理方法の説明に供するフローチャートである。また、この実施の形態では、図12に示す「依頼書」の表を処理対象として、この「依頼書」のうちの特定項目「金額」のデータである「15,000」円を特定データとして抽出する例について説明する。尚、図12に示す表においては、「金額」という特定項目は、3行1列目の矩形枠に記載されている。また、「15,000」という特定データは、3行2列目の矩形枠に記載されている。

【0205】第2の実施の形態の表処理方法によれば、表の特定項目に対応するデータ(以下、「特定データ」とも称する。)を当該表のイメージ画像から抽出するにあたり、先ず、表様式格納部48に、表のタイトルと表様式とを対応づけて表様式情報として予め格納してお

く。さらに、この表様式格納部48に、表様式毎に、当該表様式に属する項目の記載された項目枠と、当該項目に対応するデータが記載されたデータ枠との配置関係を枠配置関係情報として予め格納しておく。

【0206】このように、第2の実施の形態では、枠配置関係情報として、項目枠とデータ枠との相対的な配置関係を登録している。すなわち、特定矩形枠の表中での配置は、1箇所に限定されない。このため、1つの表様式での特定項目枠と特定矩形枠との配置関係は、枠配置関係情報の示す相対的な配置関係を満たす範囲内で任意の配置とすることができる。

【0207】ここで、図16の(A)および(B)に、それぞれ「依頼書」の表のイメージ画像の他の例を示す。図16の(A)および(B)に示す表では、図12に示す表に対して、特定データの配置が他のデータ枠と入れ替わっている。すなわち、図16の(A)に示す表では、「金額」という特定項目は、3行1列目ではなく、2行1列目の矩形枠(特定項目枠)に記載されている。また、「15,000」という特定データは、2行2列目の矩形枠(特定矩形枠)に記載されている。一方、図16の(B)に示す表では、「金額」という特定項目は、1行1列目の矩形枠(特定項目枠)に記載されている。また、「15,000」という特定データは、1行2列目の矩形枠(特定矩形枠)に記載されている。

【0208】そして、図16の(A)および(B)のいずれの表においても、特定矩形枠は、特定項目枠の右側に記載されている。従って、図16の(A)および(B)の表は、いずれも、「項目枠の右側にデータ枠がある」という相対的な配置関係を満たしている。このため、第2の実施の形態では、図12に示した表の他に、例えば図16の(A)および(B)に示した表についても、図14に示した枠配置関係情報を用いて、特定データの抽出を行うことができる。従って、第2の実施の形態によれば、特定データの配置が他のデータ枠と入れ替わっても、特定矩形枠の配置を1箇所に限定した場合に処理対象となる表の種類の数よりも、より多くの種類の表を処理対象とすることができる。

【0209】(ステップ1)そして、第2の実施の形態では、先ず、表領域・枠位置抽出部16によって、イメージ画像から、表領域を抽出すると共に、該表領域を構成する各矩形枠の位置を枠位置情報として抽出する(図15のS1)。尚、第2の実施の形態のステップ1の処理は、上述した第1の実施の形態におけるステップ1の処理と同一であるので、その詳細な説明を省略する。

【0210】(ステップ2)次に、タイトル文字列領域抽出部18および文字認識部52によって、イメージ画像のうちの非表領域から表のタイトルを抽出タイトルとして抽出する(図15のS2)。この実施の形態では、文字認識の結果、「依頼書」という抽出タイトルを抽出したとする。尚、第2の実施の形態のステップ2の処理

は、上述した第1の実施の形態におけるステップ2の処理と同一であるので、その詳細な説明を省略する。

【0211】(ステップ3)次に、表様式選択部50によって、表様式格納部から表様式情報を読み出して、当該表様式情報の中から抽出タイトルに対応する表様式を選択表様式として選択し、かつ、表様式格納部から当該選択表様式に対応する枠配置関係情報を選択枠配置関係情報として読み出す(図15のS3)。

【0212】ここで、図17を参照して、選択表様式を選択して選択表様式の枠配置関係情報を読み出す処理ステップ(S3)について説明する。図17は、選択表様式の選択処理の説明に供するフローチャートである。

【0213】この実施の形態では、選択表様式を選択するにあたり、まず、表様式選択部50が、タイトル文字列領域抽出部18から抽出タイトルを取得する(図17のS9)。ここでは、抽出タイトルとして、「依頼書」の文字コードを取得する。

【0214】次に、表様式選択部50は、表様式格納部48から、表様式情報を1つずつ順次を読み出す(図17のS10)。この実施の形態では、前述したように、表様式格納部48に、図14に示した表様式情報を格納している。ここでは、まず、1番目の表様式情報として、「注文書-様式1」を読み出す。

【0215】次に、表様式選択部50は、抽出タイトルと表様式情報とを照合する(図17のS11)。

【0216】ここでは、まず、抽出タイトルである「依頼書」の文字コードと、1回目に読み出された表様式情報のタイトルである「注文書」の文字コードとを照合する。この場合、1回目の照合結果は一致しない。

【0217】照合結果が一致しない場合(S11のステップの「no」の場合)、表様式選択部50は、次の表様式情報へアクセスして(図17のS12)、2番目の表様式情報として「依頼書-様式2」を読み出す(図17のS10)。そして、1回目と同様に、抽出タイトルである「依頼書」の文字コードと、2回目に読み出された表様式のタイトルである「依頼書」の文字コードとを照合する。この場合、2回目の照合結果は一致する。

【0218】次に、表様式選択部50は、表様式を決定する(図17のS13)。この実施の形態では、表様式選択部50は、「依頼書」に対応する「様式2」を選択表様式として決定する(図17のS13)。

【0219】この実施の形態では、上述した、表様式選択処理の照合処理(S11)は、全ての表様式情報について同様に行われる。そして、選択表様式が決定される度に、制御部(不図示)が、表様式格納部48に格納されている全表様式数を参照して、全ての表様式について照合が行われたか否かを判定する(図17のS14)。そして、全ての表様式について処理が行われていないと判定した場合(S14の「no」の場合)には、次の表様式情報にアクセスして(図17のS15)、表様式選

択部50にその表様式情報を読み出させる。そして、再び表様式情報と抽出タイトルとの照合処理(S11)を行う。

【0220】尚、この発明では、全ての表処理情報についての照合は必ずしも必要ではなく、例えば、選択表様式を決定したステップ(S13)の後直ちに次のステップ(図17のS16)へ進んで枠配置関係情報を読み出しても良い。

【0221】また、制御部(不図示)が、全ての表様式について処理が行われたと判定した場合(S14の「yes」の場合)には、表様式選択部50は、表様式格納部48から、選択表様式に対応する枠配置関係情報を読み出す(図17のS16)。この実施の形態では、「様式2」に対応する「項目枠の右側にデータ枠がある」という枠配置関係情報を読み出す。そして、図17に示す表様式選択処理を終了する。

【0222】(ステップ4)次に、表処理部54によって、枠位置情報の示す矩形枠の中から項目枠を選択する(図15のS4)。

【0223】ここで、図18を参照して、項目枠を選択する処理ステップ(図15のS4)について説明する。図18は、項目枠選択処理の説明に供するフローチャートである。この実施の形態では、表領域の矩形枠のうち、項目枠の面積は、データ枠の面積よりも一般に小さい傾向があることを利用して項目枠を選択する。

【0224】そこで、この実施の形態では、項目枠を選択するにあたり、表領域・枠位置抽出部16から枠位置情報を取得した表処理部54は、まず、矩形枠の面積を求める(図18のS17)。

【0225】次に、表処理部54は、矩形枠の面積が、一定面積以下であるかを判定する(図18のS18)。この一定面積は、任意好適な値を設定すると良い。設定にあたっては、例えば、キーボードなどの入力手段を用いて表処理を開始する前に、もしくは処理対象の表領域毎に一定面積の値を入力すると良い。

【0226】そして、判定の結果、矩形枠の面積が一定面積以下でないと判定された場合(S18の「no」の場合)は、次の矩形枠にアクセスして(図18のS19)、再び矩形枠の面積を求めて(S17)、その面積が一定面積以下であるか否かを判定する(S18)。

【0227】そして、判定の結果、矩形枠の面積が一定面積以下であると判定された場合(S18の「yes」の場合)は、表処理装置54は、その矩形枠を項目枠として選択する(図18のS20)。この実施の形態では、図12に示す表を構成する各矩形枠のうち、1行1列目の「氏名」の記載された矩形枠、2行1列目の「住所」の記載された矩形枠、3行1列目の「金額」の記載された矩形枠が項目枠として選択される。

【0228】この実施の形態では、上述した、項目枠選択処理の判定処理(S18)は、全ての矩形枠について

同様に行われる。そして、項目枠が選択される度に、制御部（不図示）が、表領域・枠位置抽出部16で抽出された全矩形枠数を参照して、全ての矩形枠について判定が行われたか否かを判定する（図18のS22）。そして、全ての矩形枠について判定処理が行われていないと判定した場合（S22の「no」の場合）には、次の矩形枠にアクセスして（図18のS23）、次の矩形枠について判定処理（S18）を行う。

【0229】また、制御部（不図示）が、全ての矩形枠について判定処理が行われたと判定した場合（S22の「yes」の場合）には、表処理部54は、項目枠の選択処理を終了する。

【0230】（ステップ5）そして、表処理部54および文字認識部52によって、項目枠の各々に記載された項目をそれぞれ抽出項目として抽出する（図15のS5）。

【0231】ここで、図19を参照して、抽出項目を抽出する処理ステップ（S5）について説明する。図19は、抽出項目を抽出する処理の説明に供するフローチャートである。

【0232】この実施の形態では、抽出項目を抽出するにあたり、表処理部54は、まず、選択された各項目枠からそれぞれ項目文字列領域を抽出する（図19のS24）。項目文字列領域の抽出にあたっては、従来公知に技術を用いることができる。例えば、表領域・枠位置抽出部16で抽出した矩形枠の内部のイメージ画像部分の周辺分布を作成して、その周辺分布における黒画素の連結成分に基づいて項目文字列領域を検出すると良い。

【0233】続いて、表処理部54は、抽出した項目文字列領域に外接する外接矩形枠の座標を文字認識部52へ転送する。尚、外接矩形枠の座標は、外接矩形枠の4頂点のx-y座標として表すと良い。

【0234】次に、文字認識部52は、外接矩形枠の座標に対応するイメージ画像について文字認識を行う（図19のS25）。文字認識部52は、文字認識にあたり、文字認識部52の内部メモリ（不図示）に記憶されている辞書を参照する。そして、文字認識部52は、辞書に格納されている標準文字のうち、認識対象と一致する標準文字の文字コードを認識結果として出力する。そして、文字認識部52は、認識結果としての文字コードを表処理部54へ転送する。この実施の形態では、認識結果として、「氏名」、「住所」および「金額」の文字コードがそれぞれ表処理部54へ転送される。

【0235】（ステップ6）次に、表処理部54によって、特定項目格納部14から特定項目を読み出して、当該特定項目が抽出項目として記載された項目枠を項目枠の中から特定項目枠として選択する（図15のS6）。

【0236】ここで、図20を参照して、特定項目枠を選択する処理ステップ（S6）について説明する。図20は、特定項目枠を選択する処理の説明に供するフロー

チャートである。

【0237】この実施の形態では、特定項目枠を選択するにあたり、まず、表処理部54が、特定項目格納部14から特定項目を読み出す（図20のS26）。ここでは、特定項目として「金額」が読み出される。

【0238】次に、表処理部54は、当該表処理部54の内部メモリ（不図示）からステップ5で選択された抽出項目を1つずつ読み出す（図20のS27）。ここでは、まず、「氏名」が読み出される。

【0239】次に、表処理部54は、特定項目と抽出項目とを比較する（図20のS28）。ここでは、まず、特定項目「金額」と1つ目の抽出項目「氏名」とを照合する。

【0240】そして、照合の結果、特定項目と抽出項目とが一致しない場合（S28の「no」の場合）には、次の抽出項目にアクセスして（図20のS29）、再び照合処理（S28）を行う。ここでは、特定項目「金額」と1番目の抽出項目「氏名」とは一致しない。このため、次の2番目の抽出項目「住所」について照合処理（S28）を行う。この場合、2番目の抽出項目「住所」についても一致しない。このため、3番目の抽出項目「金額」について照合処理（S28）を行う。

【0241】そして、照合の結果、特定項目と抽出項目とが一致した場合（S28の「yes」の場合）には、表処理部54は、その一致した抽出項目の記載された項目枠を特定項目枠として決定する（図20のS30）。ここでは、3番目の抽出項目「金額」と特定項目「金額」とが一致する。従って、表処理部54は、3番目の抽出項目「金額」の記載された3行1列目の項目枠を特定項目枠として選択する。

【0242】この実施の形態では、上述した、特定項目と抽出項目との照合処理（S28）は、全ての抽出項目について同様に行われる。そして、特定項目枠が決定される度に、制御部（不図示）が、表処理部54の内部メモリに格納されている全抽出項目数を参照して、全ての抽出項目について照合が行われたか否かを判定する（図20のS31）。そして、全ての抽出項目について処理が行われていないと判定した場合（S31の「no」の場合）には、次の抽出項目にアクセスして（図20のS32）、次の抽出項目について照合処理（S28）を行う。

【0243】尚、この発明では、全ての項目についての照合は必ずしも必要ではなく、例えば、特定項目枠を決定したステップ（S30）の後直ちに特定項目枠選択処理を終了しても良い。

【0244】また、制御部（不図示）が、全ての項目について処理が行われたと判定した場合（S31の「yes」の場合）には、特定項目枠選択処理を終了する。ここでは、「金額」の記載された3番目の抽出項目の照合処理を行うことで、全ての抽出項目について照合処理が

行われたことになるので、S32のステップに進むことなく特定項目枠の選択処理を終了する。

【0245】(ステップ7)次に、表処理部54によって、特定項目枠の位置に対して選択枠配置関係情報の示す位置にある矩形枠をイメージ画像から特定矩形枠として抽出する(図15のS7)。

【0246】ここで、図21を参照して、特定矩形枠を抽出する処理ステップ(S7)について説明する。図21は、特定矩形枠を抽出する処理の説明に供するフローチャートである。

【0247】この実施の形態では、特定矩形枠を抽出するにあたり、まず、表様式選択部50から選択表様式に対応する枠配置関係情報を取得する(図21のS33)。ここでは、「様式2」に対応する、「項目枠の右側に矩形枠がある」という枠配置関係情報を取得する。

【0248】次に、表処理部54は、表処理部54の内部メモリに記憶されていた特定項目枠を取得する(図21のS34)。この実施の形態では、特定項目枠として「3行1列目」を取得する。

【0249】次に、表処理部54は、この特定項目枠に対して、枠配置関係情報の示す位置にある矩形枠を特定矩形枠として抽出する(図21のS35)。この実施の形態では、特定項目枠「3行1列目」に対して、枠配置関係情報「項目枠の右側にデータ枠がある」という関係にある矩形枠、すなわち、「3行2列目」の矩形枠を特定矩形枠として抽出する。

【0250】例えば、特定項目枠が複数個選択されている場合、上述した特定矩形枠の抽出は、全ての特定項目枠について行われる。そして、特定矩形枠が抽出される度に、制御部(不図示)が、表処理部54の内部メモリに格納されている全特定項目枠数を参照して、全ての特定項目枠について処理(S35)が行われたか否かを判定する(図21のS36)。そして、全ての特定項目枠について処理(S35)が行われていないと判定した場合(S36の「no」の場合)には、次の特定項目枠にアクセスして(図21のS37)、その特定項目枠についての特定矩形枠を抽出する(S35)。

【0251】また、制御部(不図示)が、全ての特定項目枠について処理(S35)が行われたと判定した場合(S36の「yes」の場合)には、特定矩形枠の抽出処理を終了する。この実施の形態では、特定項目枠は1つであるので、特定矩形枠を抽出した後、S37へ進むことなく特定矩形枠抽出処理を終了する。

【0252】(ステップ8)次に、表処理部54によって、特定矩形枠内から特定データを抽出する(図15のS8)。この実施の形態では、「15,000」という金額の特定データを抽出する。尚、第2の実施の形態のステップ8の処理は、上述した第1の実施の形態におけるステップ6の処理と同一であるので、その詳細な説明を省略する。

【0253】続いて、表処理部54は、この「15,000」の文字コードを特定データとして表処理装置46の外部へ出力する。

【0254】このように、第2の実施の形態の表処理方法によれば、表領域の中の項目枠についてのみ文字認識を行うことにより、特定データを選択的に抽出できる。その結果、この表処理方法によれば、表領域の全てについて文字認識を行う場合に要する処理時間に比べて、表処理に要する処理時間の短縮を図ることができる。

【0255】

【発明の効果】第1の表処理方法および装置によれば、タイトルと表様式を対応づけて登録して格納しておく。このため、イメージ画像から抽出されたタイトルを用いて、表様式を決定することができる。また、各表様式について、当該表様式に属する項目毎のデータ枠の配置をそれぞれ登録して格納しておく。このため、タイトルを抽出し、かつ特定項目を指定することにより、特定項目に対応するデータ枠の表中に配置(特定枠配置)を決定することができる。すなわち、表領域について文字認識を行う前に、特定枠配置を決定することができる。その結果、イメージ画像のうち、特定枠配置に相当する矩形枠の領域について選択的に文字認識をすることによって、特定データを抽出することができる。すなわち、表領域の全てについて文字認識を行わなくとも、特定データを抽出することができる。このため、表中の全ての文字パタンの文字認識を行う場合に比べて、短時間で特定データを抽出することができる。その結果、表処理に要する時間の短縮を図ることができる。

【0256】また、第2の表処理方法および装置においては、タイトルと表様式とを対応づけて表様式情報として予め登録しておく。このように登録しておけば、イメージ画像から抽出したタイトル(抽出タイトル)を用いて、表様式を選択することができる。さらに、この処理方法では、表様式毎に、項目枠とデータ枠との配置関係を枠配置関係情報として予め登録しておく。このように登録しておけば、表様式を選択することにより、その選択表様式における項目枠とデータ枠との配置関係(例えば、その選択表様式においては、項目枠の右側に隣接してデータ枠が配置されているという配置関係)を定めることができる。

【0257】一方、第2の表処理方法および装置においては、表領域の矩形枠の中から項目枠を選択する。続いて、選択された項目枠について文字認識を行って、各項目枠から項目を抽出する。そして、項目枠の中から特定項目が記載された特定項目枠を選択する。特定項目は、例えば予め指定しておくとも良い。

【0258】従って、この表処理方法においては、選択された特定項目枠の位置に対して、枠配置関係情報の示す配置関係に位置するデータ枠を特定矩形枠として選択することができる。すなわち、表領域の矩形枠のうち、

項目枠についてのみ文字認識を行うことにより、特定データを選択的に抽出できる。その結果、表領域の全てについて文字認識を行わなくとも、特定データを選択的に抽出することができる。このため、表領域の全てについて文字認識を行ってから特定データを抽出する場合に要する処理時間に比べて、この処理方法では、特定データの抽出に要する処理時間の短縮を図ることができる。すなわち、表処理に要する時間の短縮を図ることができる。

【0259】さらに、第2の表処理方法および装置によれば、枠配置関係情報として、項目枠とデータ枠との相対的な配置関係を登録している。すなわち、特定矩形枠の表中での配置は、1箇所に限定されない。このため、1つの表様式での特定矩形枠の配置は、枠配置関係情報の示す相対的な配置関係を満たす範囲内で任意の配置とすることができる。従って、第2の表処理方法によれば、特定矩形枠の配置を1箇所に限定した場合に処理対象となる表の種類の数よりも、より多くの種類の表を処理対象とすることができる。

【図面の簡単な説明】

【図1】第1の実施の形態の表処理装置の構成の説明に供する機能ブロック図である。

【図2】表様式情報および枠配置情報の例を示す図である。

【図3】表領域・枠位置抽出部の説明に供する機能ブロック図である。

【図4】表様式・枠配置選択部の説明に供する機能ブロック図である。

【図5】第1の実施の形態の表処理方法の説明に供するフローチャートである。

【図6】表領域・枠位置抽出処理の説明に供するフローチャートである。

【図7】抽出タイトル抽出処理の説明に供するフローチャートである。

【図8】選択表様式選択処理の説明に供するフローチャートである。

【図9】特定枠配置選択処理の説明に供するフローチャートである。

【図10】特定矩形枠選択処理の説明に供するフローチャートである。

【図11】特定データ抽出処理のフローチャートである。

【図12】処理対象の表のイメージ画像例を示す図である。

【図13】第2の実施の形態の表処理装置の構成の説明に供する機能ブロック図である。

【図14】表様式情報および枠配置関係情報の例を示す図である。

【図15】第2の実施の形態の表処理方法の説明に供するフローチャートである。

【図16】(A)および(B)は、処理対象の表のイメージ画像例を示す図である。

【図17】表様式選択処理の説明に供するフローチャートである。

【図18】項目枠選択処理の説明に供するフローチャートである。

【図19】項目抽出処理の説明に供するフローチャートである。

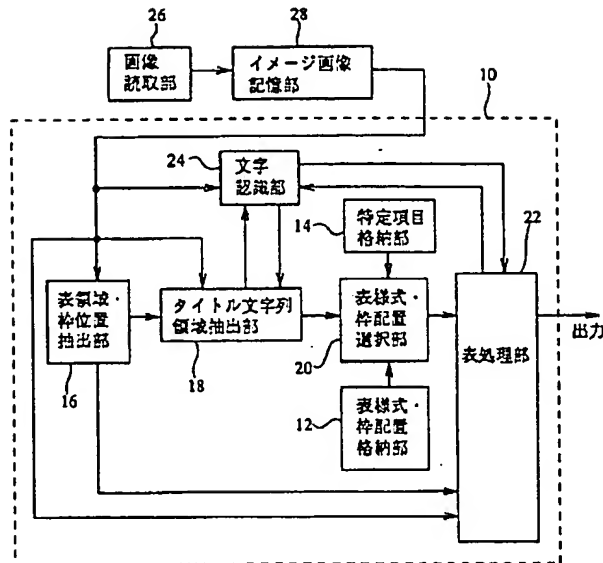
【図20】特定項目枠選択処理の説明に供するフローチャートである。

【図21】特定矩形枠抽出処理の説明に供するフローチャートである。

【符号の説明】

- 10：表処理装置
- 12：表様式・枠配置格納部
- 14：特定項目格納部
- 16：表領域・枠位置抽出部
- 18：タイトル文字列領域抽出部
- 20：表様式・枠配置選択部
- 22：表処理部
- 24：文字認識部
- 26：画像読取部
- 28：イメージ画像記憶部
- 30：周辺分布作成部
- 32：罫線抽出部
- 34：枠位置抽出部
- 36：表領域抽出部
- 38：タイトル照合部
- 40：表様式選択部
- 42：項目照合部
- 44：特定枠配置選択部
- 46：表処理装置
- 48：表様式格納部
- 50：表様式選択部
- 52：文字認識部
- 54：表処理部
- 58：表領域
- 60：特定矩形枠

【図1】



10: 表処理装置

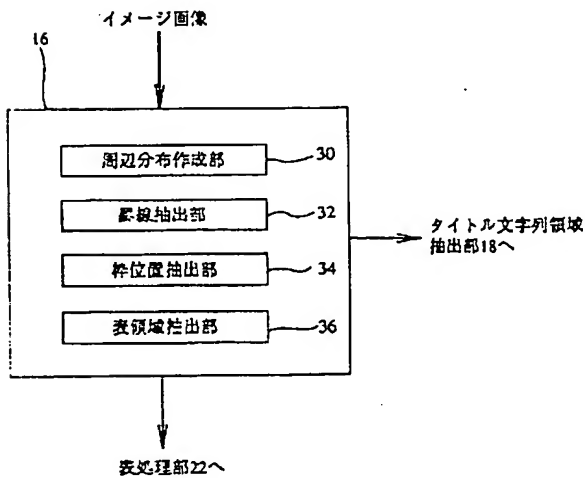
第1の実施の形態

【図2】

タイトル	表様式	項目	データ枠の配置
注文書	様式1	氏名	第1行第2列
		品名	第2行第2列
依頼書	様式2	氏名	第1行第2列
		住所	第2行第2列
		金額	第3行第2列
納品書	様式3	品名	第2行第1列
		金額	第2行第2列

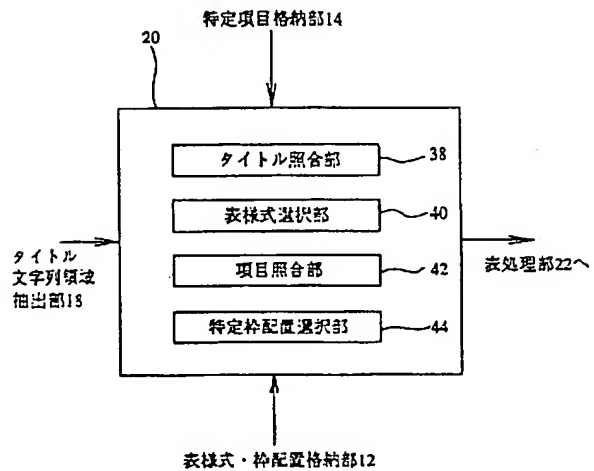
表様式情報および枠配置情報の例

【図3】



表領域・枠位置抽出部の機能ブロック図

【図4】



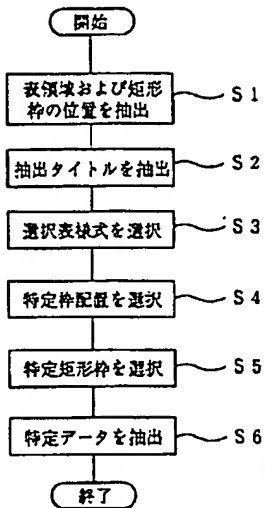
表様式・枠配置選択部の機能ブロック図

【図14】

タイトル	表様式	枠配置関係
注文書	様式1	項目枠の下側にデータ枠がある
依頼書	様式2	項目枠の右側にデータ枠がある

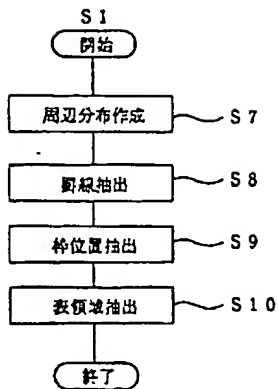
表様式情報および枠配置関係情報の例

【図5】

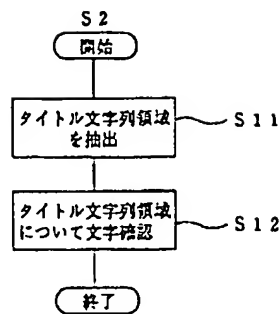


表領域・枠位置抽出処理のフローチャート

【図6】



【図7】



抽出タイトル抽出処理のフローチャート

【図12】

依頼書

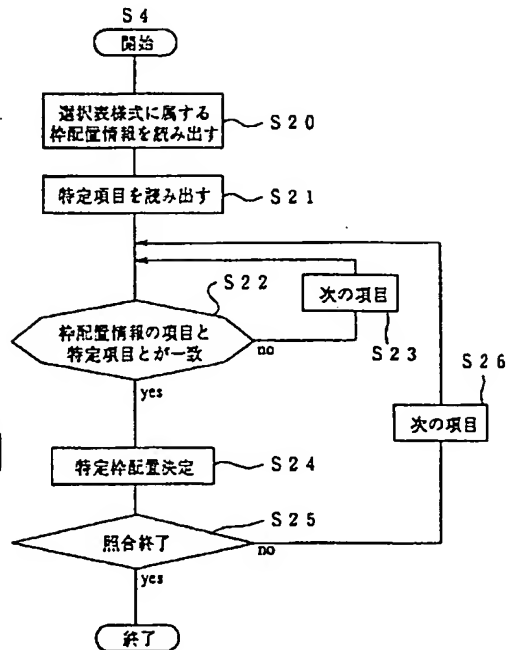
氏名	○×△
住所	△市○町1-1
金額(円)	15,000

58: 表領域 60: 特定矩形枠

表のイメージ図案例

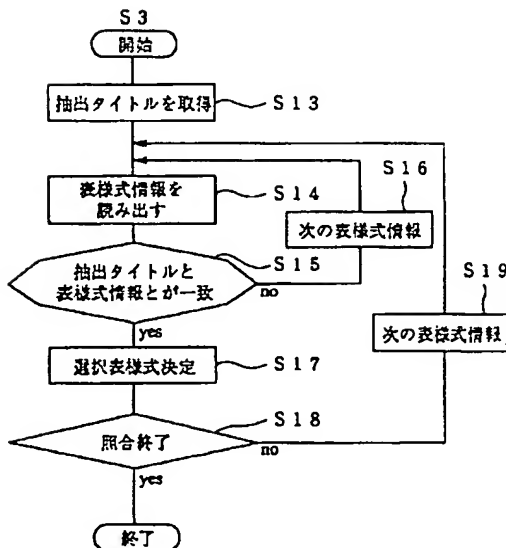
第1の実施の形態のフローチャート

【図9】



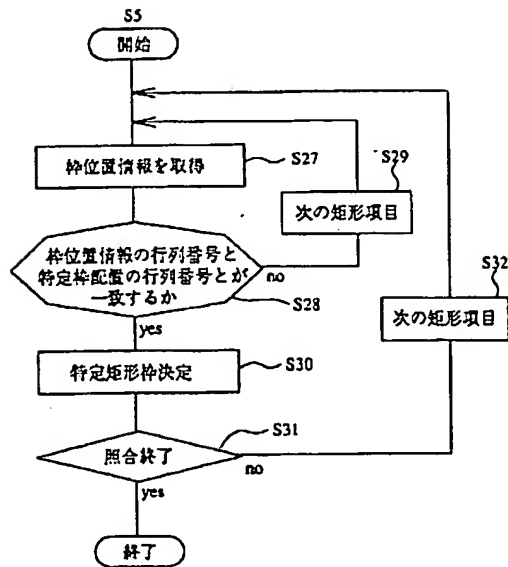
特定枠配置選択処理のフローチャート

【図8】



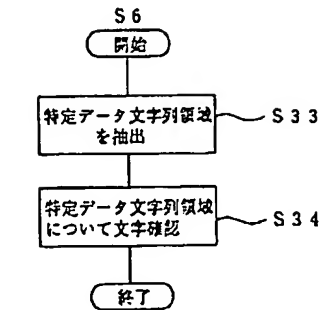
選択表様式選択処理のフローチャート

【図10】



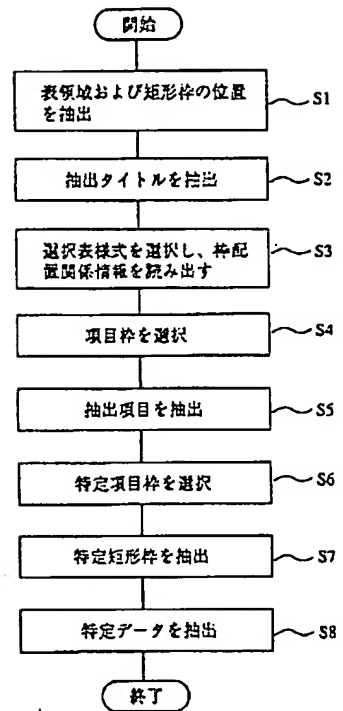
特定矩形枠選択処理のフローチャート

【図11】



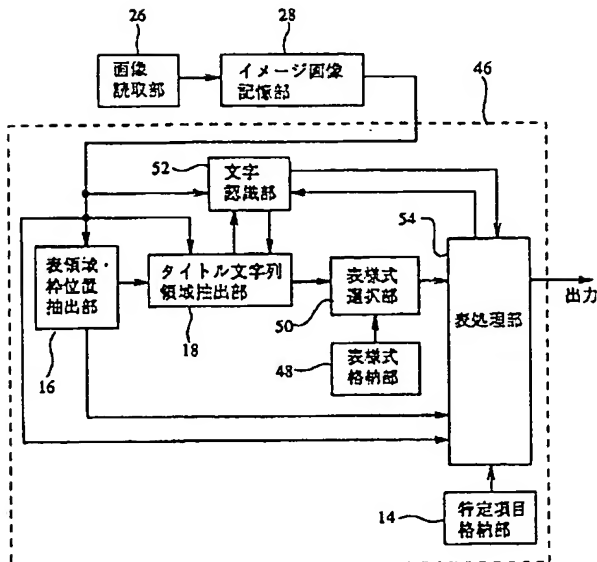
特定データ抽出処理のフローチャート

【図15】



第2の実施の形態のフローチャート

【図13】



46: 表処理装置

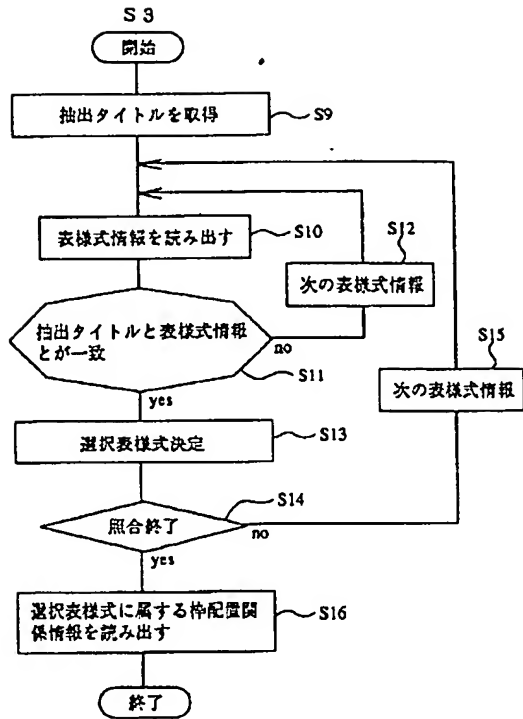
第2の実施の形態

【図16】

依頼書							
(A)	<table border="1"> <tr><td>氏名</td><td>〇×△</td></tr> <tr><td>金額 (円)</td><td>15,000</td></tr> <tr><td>住所</td><td>△市〇町1-1</td></tr> </table>	氏名	〇×△	金額 (円)	15,000	住所	△市〇町1-1
氏名	〇×△						
金額 (円)	15,000						
住所	△市〇町1-1						
依頼書							
(B)	<table border="1"> <tr><td>金額 (円)</td><td>15,000</td></tr> <tr><td>氏名</td><td>〇×△</td></tr> <tr><td>住所</td><td>△市〇町1-1</td></tr> </table>	金額 (円)	15,000	氏名	〇×△	住所	△市〇町1-1
金額 (円)	15,000						
氏名	〇×△						
住所	△市〇町1-1						

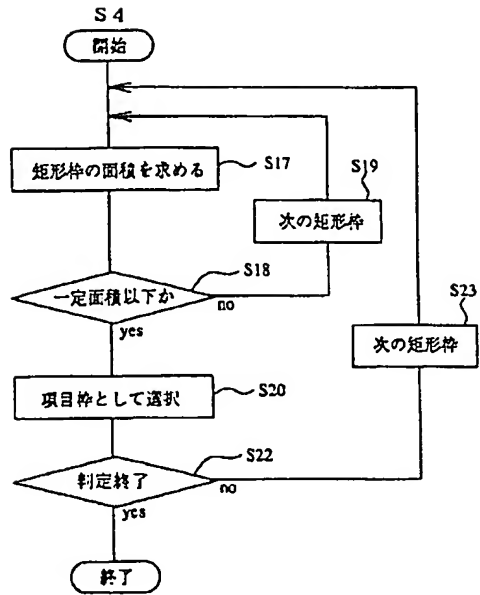
表のイメージ画像列

【図17】



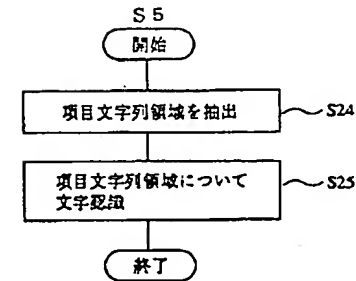
選択表様式選択処理のフローチャート

【図18】



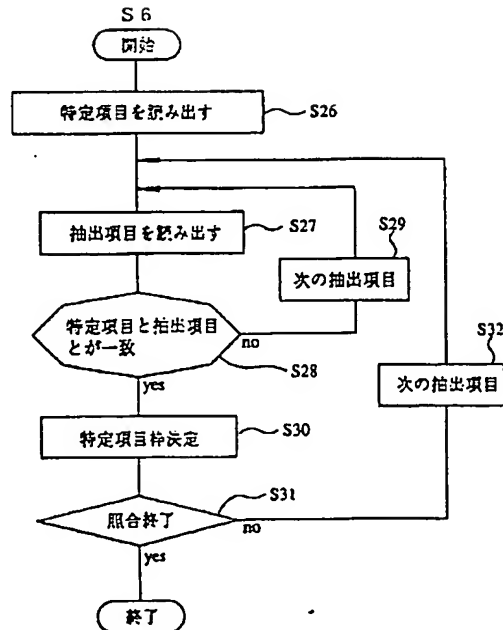
項目枠選択処理のフローチャート

【図19】



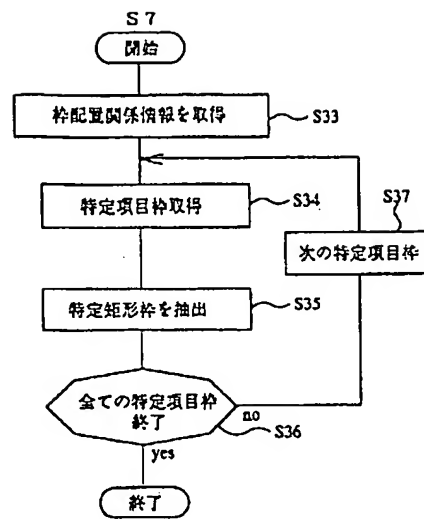
抽出項目抽出処理のフローチャート

【図20】



特定項目枠選択処理のフローチャート

【図21】



特定矩形枠抽出処理のフローチャート